[D1]
THE CHAIN LADDER TECHNIQUE — A STOCHASTIC MODEL
Contributed by B Zehnwirth

1. Introduction

The chain ladder technique (equivalently, age-to-age development factors) is one of the oldest actuarial techniques to be applied widely for estimating loss reserves.

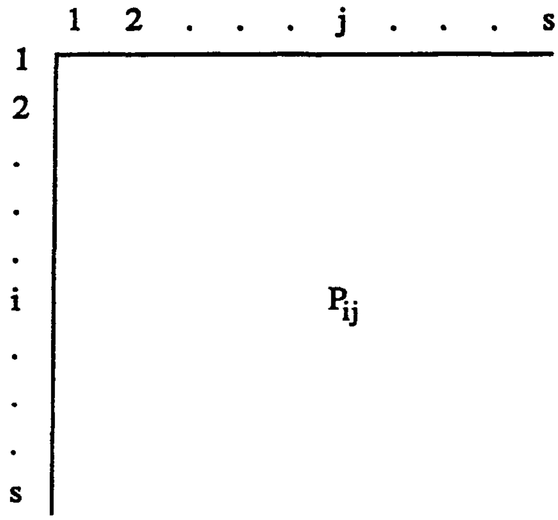
The technique appears intuitively natural and only until more recently was always regarded as being based on a non-stochastic model: that is, a model which is deterministic and accordingly does not include a random component.

The principal objective of this article is to demonstrate the intimate connection between the chain ladder technique and a two-way analysis of variance model applied to the logarithms of the incremental paid losses. Recognition of this connection reveals the merits and defects of the chain ladder technique more clearly.

2. Chain ladder technique

We first review the chain ladder technique in order to indicate two underlying model assumptions. The second model assumption is often not recognised by many users of the technique.

Let P_{ij} represent the incremental paid loss made in development year j, in respect of accident year i. The batch of data P_{ij}, i=1,...,s; j=1,...,s−i+1 is represented as a matrix thus:



Accident years (rows) range from 1 to s and development years (columns) also range from 1 to s.

We denote the cumulative paid loss in development year j, in respect of accident year I by C_{ij}. It is given by:

$$C_{ij} = \sum_{h=1}^{j} P_{ij} \; .$$

A matrix of development factors based on the { C_{ij} } array is constructed by computing the development factor D_{ij} as

$$D_{ij} = \frac{C_{ij}}{C_{i\,j-1}} \qquad i = 1,\ldots,s; \quad j = 2,\ldots,s-i+1$$

The first basic assumption made is

Assumption 1: Each accident year has the same age-to-age development factors. Equivalently, for each j=2,...,s

$$D_{ij} = D_j \qquad \text{for all } i = 1,2,\ldots s.$$

Under Assumption 1, the most popular estimator of the development factor D_j is the weighted average

$$\hat{D}_j = \frac{\sum_{i=1}^{s-j+1} C_{ij}}{\sum_{i=1}^{s-j+1} C_{i\,j-1}}$$

$$= \frac{\sum_{i=1}^{s-j+1} C_{i\,j-1} * D_{ij}}{\sum_{i=1}^{s-j+1} C_{i\,j-1}}$$

The development factor D_{ij} is weighted by the corresponding "volume" measure $C_{i\,j-1}$.

Some users of the chain ladder technique do not use the weighted average estimator of D_j. This is an estimation issue that we address subsequently in this chapter. The fact remains that Assumption 1 is a model assumption associated with the chain ladder technique.

Projections of the quantities C_{ij}; i=2,...,s; j=s−i+2,...s are computed thus:

$$\hat{C}_{ij} = C_{i\ s-i+1} \prod_{k=s-1+2}^{s} \hat{D}_k$$

This technique of projection is explicitly based on the fact that a second model assumption is valid. It is assumed that each accident year has **necessarily** a different level estimated by that year's individual experience. The quantity $C_{i\ s-i+1}$ represents an estimate of the level of accident year i.

Assumption 2: Each accident year has a parameter representing its level. The level parameter for accident year I is estimated by $C_{i\ s-i+1}$.

The last accident year s is represented by the single observation C_{i1}. Were we to assume that accident years are **completely** homogeneous, we should estimate the level of accident years by

$$\sum_{i=1}^{s} C_{i1}/s\ ,$$

(or a better estimator of the mean level at development year 1).

Complete homogeneity means that the observations C_{i1}, C_{21},...,C_{s1} are generated by the same mechanism. The chain ladder technique explicitly assumes that the mechanisms generating the incremental paid losses C_{i1}, C_{21},...,C_{s1} are so unrelated that pooling of the information does not afford any increased efficiency. I would find it very difficult to believe that this assumption is **ever** true. In any case, why not find out first what the data indicate?

3. Statistical models related to chain ladder technique

Based on the two assumptions discussed in the preceding section, the following autoregressive model discussed in the paper by Kamreiter and Straub (1973) suggests itself.

$$C_{ij} = D_j\, C_{i\,j-1} + \delta_{ij}\ ;\ i = 1,...,s$$

where the random variables D_j, δ_{ij} and $C_{i\,j-1}$ are independent and satisfy

$$E[\delta_{ij}] = 0,\ E[D_j] = d_j.$$

The quantities $\{ D_j \}$ represent the development factors and are the same for each accident year. Note that it is implicitly assumed that the observations C_{i1}, C_{21},..., C_{s1} are **not** related (at all). Moreover, the additive error term δ_{ij} is questionable — the error term should be multiplicative (see Section 4).

We only remark that the above mentioned model satisfies Assumptions 1 and 2 of the preceding section and devote the remainder of this chapter to a second stochastic model, discussed by Kremer (1982).

The basic model is defined by the multiplicative representation,

$$P_{ij} = a_i' \, . \, b_j' \, e_{ij}' \tag{2.1}$$

where a_i' is the parameter representing the effect of accident year I;
b_j' is the parameter representing the effect of development year j;

and e_{ij}' is a random error term.

By taking logarithms of both sides of equation (2.1), the model may be re-formulated as a two-way analysis of variance model, viz.,

$$Y_{ij} = \log P_{ij} = \mu + a_i + b_j + e_{ij} \tag{2.2}$$

where the parameter μ represents the overall mean effect (on a logarithmic scale), the parameter a_i represents the residual effect due to accident year i and the parameter b_j represents the residual effect due to development year j. It is also assumed that

$$\sum_{i=1}^{s} a_i = \sum_{i=1}^{s} b_j = 0 \tag{2.3}$$

and that $\{ e_{ij} \}$ represent zero mean uncorrelated errors with $\text{Var}[e_{ij}]=\sigma^2$.

This model implies that each incremental paid loss P_{ij} has a lognormal distribution.

In the two-way analysis of variance model (2.2), accident year is regarded as a factor at s levels and development year is regarded as a factor at s levels. It is also assumed that the P_{ij}'s are independent random variables having a lognormal distribution with

$$\text{mean} = \exp(\mu + a_i + b_j + 0.5\sigma^2) \tag{2.4}$$

and

$$\text{variance} = \text{mean}^2 * (\exp(\sigma^2) - 1) \tag{2.5}$$

Accident year effects and development year effects are assumed to be additive with no interaction. In other words, the effect of an accident year is the same for each development year and vice versa.

We now turn to the estimation of the parameters μ, $\{ a_i \}$ and $\{ b_j \}$.

Model (2.2) is essentially a regression model where the design matrix involves indicator variables. However, the design based on (2.2) alone is singular. In view of constraint (2,3), the actual number of free parameters is $2s-1$, yet model (2.2) has $2s+1$ parameters. By setting $a_1=b_1=0$, say, the resulting design is non-singular and estimates of parameters can be obtained using a statistical regression package.

Kremer (1982) presents three recursive equations for estimating the parameters μ, $\{ a_i \}$ and $\{ b_j \}$. These equations are essentially solutions to the normal equations of the model described by expression (2.2) and constraint (2.3). If there are no missing data values in the matrix, estimates of the parameters can be obtained using standard methods. When there are too many missing values, and standard methods cannot be used, the following technique, called the E–M algorithm has a fair amount of intuitive appeal.

For a complete matrix the estimates of the parameters are well known:

$$\hat{\mu} = \bar{Y}_{..} = \sum_{i=1}^{s} \sum_{j=1}^{s} Y_{ij}/s^2, \tag{2.6}$$

$$\hat{a}_i = \bar{Y}_{i.} - \bar{Y}_{..} , \tag{2.7}$$

and

$$\hat{b}_j = \bar{Y}_{.j} - \bar{Y}_{..} , \tag{2.8}$$

where

$$\bar{Y}_{i.} = \sum_{j=1}^{s} Y_{ij}/s \tag{2.9}$$

$$\bar{Y}_{.j} = \sum_{j=1}^{s} Y_{ij}/s \tag{2.10}$$

The E–M algorithm is a recursive technique for finding maximum likelihood estimates in the case of incomplete data. The estimates given by (2.6) to 2.8) are maximum likelihood but are based on a complete matrix. The E in the term "E–M algorithm" stands for Expectation and the M for Maximisation (of the likelihood).

Step 0: Complete the matrix by starting with some initial expected values. For instance, you may enter into the (empty) cell (i,j) the value $y_{i,s-i+1}$.

Step 1: Compute the maximum likelihood estimates for the completed matrix using equations (2.6) to (2.8).

Step 2: Use the estimates $\hat{u}$, $\{\hat{a}_i\}$ and $\{\hat{b}_j\}$ obtained in Step 1 to compute new expected values $\hat{u}, + \hat{a}_i + \hat{b}_j$ for the empty cells (lower triangle).

Now return to Step 1 and continue the recursions until a certain prescribed tolerance is reached, e.g. relative change in all estimates is less than 10^{-3}.

The final estimates $\hat{u}$, $\{\hat{a}_i\}$ and $\{\hat{b}_j\}$ represent the maximum likelihood estimates. The variance σ^2 is estimated by the Mean Square Error

$$\hat{\sigma}^2 = \sum_{i=1}^{s} \sum_{j=1}^{s-i+1} (y_{ij} - \hat{\mu} - \hat{a}_i - \hat{b}_j)^2 \; / \; (n - 2s - 1)$$

where n=total number of observations in the upper triangle, viz., $s(s+1)/2$.

Forecasts of P_{ij} for $i=2,\ldots, s$ and $j=s-i+2,\ldots, s$ are given by

$$\hat{P}_{ij} = \exp\,(\hat{\mu} + \hat{a}_i + \hat{b}_j + 0.5\hat{\sigma}^2).$$

Note that the two-way analysis of variance model can be applied and estimated for **any shape** array of the **incremental paid losses**. This means that a **formal** chain ladder technique can be applied to any shape array provided $n>2s-1$.

4. The importance of the log transform — removal of heterogeneity

Loss reservers often describe their data as being heterogeneous. For a long tail line of business, payments are necessarily made over time. Indeed, the main cause of heterogeneity is time itself! Time, almost always, almost everywhere, subjects incremental paid losses (and severities) to one type of heterogeneity we already know about: the variability in incremental paid losses (and in severities) increases as mean level increases.

Let's illustrate this well supported phenomenon with an example. If in 1965 average severity was 1,000 and standard deviation of severity 200, and if in 1988 average severity is 30,000, then the standard deviation of severity in 1988 is probably around 6,000. However, the standard deviation of the logarithms of severities has remained stable between 1965 and 1988. The logarithmic transformation stabilises the variance since it has a standard deviation that is proportional to the mean.

Based on the foregoing discussion, the model

$$P_{ij} = \mu + a_i + b_j + e_{ij}$$

in place of model (2.1) of Section 3, cannot be correct because the variance of the error term e_{ij} will necessarily depend on μ, a_i and b_j.

The foregoing discussion, moreover, also indicates that the geometric mean of development factors is a more efficient estimate of the mean development factor than an arithmetic average.

5. Estimation of development factors

Development factors are typically based on the cumulative paid losses and are ratios of numbers. It is not possible to determine, by eye, if two computed development factors are different in the sense that they are generated by a different process. For example, suppose the incremental paid losses for the first two development years, for two contiguous accident years, are generated by 100 tosses of a symmetric coin. The following scenario may be observed.

		Development Year	
		0	1
Accident	1	41	63
Year	2	59	38

The two computed development factors are 2.537 and 1.644. These, however, are generated by the same process.

Moreover, there is a substantial loss of information when data are cumulated. For instance, a constant incremental paid loss of 100 at every development year has development factors based on cumulative data that asymptote to one, and indeed, even if the incremental paid losses increase according to a polynomial trend, the development factors (based on the cumulative data) asymptote to one. Furthermore, any trends in the payment year direction are different to identify and estimate if the data are cumulated in the development direction.

6. Parameters

Consider the following quadratic trend model representing annual sales of a product,

$$y_t = 2 + 3t^2 + n_t$$

where $t = 1,2,...$ denotes year, y_t sales in year t, and the error terms $\{ n_t \}$ are zero mean and independent from a Normal distribution with variance σ^2.

Suppose we generate the values $y_1, y_2, ..., y_7$ (seven year sales figures) and ask a colleague to forecast y_8. We know the complete specification of the model generating the sales including σ^2.

The colleague estimates the following models:

1. *Linear trend*

$$y_t = a_0 + a_1 t + n_t.$$

The regression output indicates that R^2=68% but the residuals appear to have a systematic pattern.

2. *Quadratic trend*

$$y_t = a_0 + a_1 t + a_2 t^2 + n_t$$

For this model R^2=76% and the residuals appear to be in good shape.

The colleague observes that as the number of parameters increases, the quality of fit is improved as measured by R^2. Accordingly, the next model suggests itself.

3. *A polynomial of degree six*

$$y_t = a_0 + a_1 t + \ldots a_6 t^6$$

Here R^2=100% and the fitted curve presents residuals that are all zero.

The colleague presents his forecast as

$$y_8 = a_0 + 8a_1 + \ldots + 8^6 a_6$$

When the colleague presents his solution, we mention to him that the data presented to him had an error, that is, the datum y_4 had been incorrectly generated.

The colleague has now to revise his forecast in the light of this information — the revised forecast is likely to bear no resemblance to the first forecast especially if σ^2 is large!!

The moral of this tale is that the polynomial model used by the colleague produces forecasts that are extremely sensitive to the random component in the data. The forecasts are subject to large uncertainties and accordingly are **not** useful. This is a feature possessed by any model that has many parameters — overparametrisation results in instability. The chain ladder model or technique has many parameters. An array comprising s accident years and s development years involves 2s−1 parameters. In particular, there is an accident year parameter for accident year s where there is only one observation — similarly for development year s.

Every model contains a priori information — the chain ladder **model** contains very little a priori information. The chain ladder model does not contain any information in respect of:

(i) trends and/or patterns in development factors;
(ii) trends across accident years;
(iii) trends across payment years.

Typically in Statistics, a two-way analysis of variance model is applied to a rectangular array involving two factors, each at a number of levels. A factor is a qualitative variable. We normally do not relate the different levels of a factor. For example, when analysing the effects of different soil types and fertilisers on yield of barley, we do not assume some kind of trend or systematic pattern across the fertilisers! It is absurd to treat accident years and development years as factors at different levels, the way we treat different soil types and different fertilisers.

The example involving the sixth degree polynomial gives us some insight as to when the chain ladder technique may work (provided the parameters are estimated efficiently). The chain ladder technique works when the mechanisms generating the paid losses are completely deterministic, that is, $\sigma^2 = 0$, or σ^2 is very close to 0 and development factors are homogeneous. Unfortunately, the real world is not like that.

References

Kramreiter, H. and Straub, E. (1973), On the calculation of IBNR reserves II, *Swiss Actuarial Journal*, 73, pp 177–190.

Kremer, E. (1982), IBNR — claims and the two-way model of ANOVA, *Scandinavian Actuarial Journal*, 1, pp 47–55.

◊