MODELLING LOSS RESERVE DEVELOPMENTS

ROBERT J. FINGER

The actuarial analysis of loss reserve developments begins by analyzing the patterns in historical claim data. Implicitly this analysis proceeds from a variety of assumptions, which may or may not be acknowledged or tested. By projecting loss reserves from historical data, the analyst is essentially using a mathematical model. This paper presents a general approach aimed at developing and exploring as many alternative models as possible. It is felt that there are indeed some patterns which will continue into the future; at times it may be extremely difficult to uncover these patterns or to even know that they exist. Looking backwards in time, however, it is always possible to describe what has occurred; the historical patterns may be erratic or largely meaningless, but they do exist. Likewise, at some distant future date it should be equally possible to describe the payout on the loss reserves which must now be estimated.

TYPICAL ASSUMPTIONS

Actuarial literature gives many examples of assumptions which are made to create mathematical models of reserve developments. Tarbell[1], for example, assumed that the incurred but not reported (IBNR) liability was proportional to the ratio of incurred losses in the last three months of the last two years. Fisher and Lange[2] assumed that the inflation rate (change in average cost per claim) remains constant for each different age of claim at settlement. Resony[3] assumes that the ratio of paid allocated loss expenses to claims disposed of (change in outstanding reserve by reported year) is constant by age of settlement. A common method of calculating loss development factors for ratemaking purposes assumes that the change in incurred losses from development period to development period will remain constant. On closer scrutiny this single assumption is a composite of several others, such as:

- the reporting pattern of claims by development period will not change and
- the degree of underreserving or overreserving will not change or
- violations of the above assumptions will exactly offset each other.

---

[1] Tarbell, T. F. "Incurred But Not Reported Claim Reserves," PCAS XX (1934).

[2] Fisher, W. H., and Lange, J. T., Loss Reserve Testing: A Report Year Approach," PCAS LX (1973).

[3] Resony, A. V. "Allocated Loss Expense Reserves," PCAS LIX (1972).

TERMINOLOGY AND DATA

This paper will refer to a liability as a fixed, though perhaps unknown, amount of money which is owed to others. The term reserve is used to mean an estimate of a liability.

Virtually all loss development data can be put into the characteristic matrix format. This matrix is shown below.

Characteristic Matrix Format

Exposure Period	Development Period 1	2	3	4
1	O_{11}	O_{12}	O_{13}	O_{14}
2	O_{21}	O_{22}	O_{23}	O_{24}
3	O_{31}	O_{32}	O_{33}	
4	O_{41}	O_{42}		
5	O_{51}			

The O_{ij} are observations of some type of reserve data for exposure period i as of development period j. Development periods are successive evaluations of loss development data. Periods are often of twelve-months durations, but can be of one, three, or six-month durations. Exposure periods are groupings of claims. Claims may be grouped by accident years, policy years, report years, or other durations. Exposure periods may also represent groups of claims which were part of a liability (such as the case reserve or IBNR reserve) at a point in time. It is typical, but not necessary, that the durations of the development and exposure periods be the same. This happens, for example, when accident year data is evaluated every 12 months.

Various types of data can fill the characteristic matrix format. The basic variations are: (1) incurred or paid losses and (2) cumulative or incremental developments. The amounts reflected can be aggregate claim amounts, claim counts, or average claim amounts. The amounts can include or exclude allocated loss adjustment expense, subrogation and salvage, ceded and assumed reinsurance, and perhaps deductibles or reinsurance retentions. Claim counts can be defined to include or exclude claims closed without an indemnity payment. Amounts can include several lines of business or coverages.

Various reserving methods utilize the characteristic matrix format in different ways. Among the five most common ways are: (1) using the entire matrix, (2) using one or more diagonals of the matrix, (3) using the ratios of one matrix to another (such as the ratio of paid allocated expense to paid losses), (4) using different matrices for different lines of business (additive combinations), or (5) using a multiplicative combination of matrices (such as those for the number of claims and average claim amounts).

TYPES OF DEVELOPMENT PATTERNS

Assumptions made by a loss reserving method are related to patterns in the characteristic matrix format. The analyst can test the previous accuracy of the assumptions by evaluating the matrix. Further, the analyst can evaluate the potential applicability of other assumptions by reviewing the matrix. In particular, there are a variety of relationships or patterns which are used in different reserving methods.

In analyzing the characteristic matrix format, vertical groups of data represent evaluations of successive groups of claims at the same stage (duration) of development. Horizontal data groups represent successive evaluations of the same group of claims. Diagonal data groups represent developments which occurred during the same calendar period of time.

Many loss reserving methods assume a consistent relationship between two variables, as expressed by the ratio between them. The use of the claim count and the average claim amount is a common example. In this case the average claim amount is actually a ratio of the aggregate losses to the claim count. Ratios may be between two different claim-related variables, between a claim and a premium or exposure variable, or between claims and an external variable. In the first case, an example is the ratio of paid loss expenses to paid losses. In the second case, loss ratios or pure premiums may be evaluated. In the latter case, inflation indices can be used.

Another possible relationship is to model loss developments by a probability distribution. The reporting or payment of claims could, for example, be modeled as a cumulative distribution function in time. A problem that arises with this approach is that time is unbounded, whereas at some point all claims will certainly be reported and all payments will be made. A possible solution would be to fix a certain time period as the ultimate development.

Another possible solution is a different way of looking at reserve developments: claim turnover intervals. Instead of assuming that the development period affects the loss development, it is assumed that the percentage of claims which have been closed affects it. For example, it is assumed that the seventieth to eightieth percentile of closed claims have a constant pattern. (Data is graphically portrayed in this format in Figure 1.) This assumption is useful for lines of business where the claims which remain open a longer period of time close at significantly higher average amounts.

THE BASIC MODEL

All of the previous data formats and relationships can be represented by a generalized loss reserve development model. This model is defined as follows:

(1) $$O_{ij} = C_{ij}F_jS_iK_{i+j} + e_{ij}$$

Where:

- O_{ij} — is the observed values of the process for exposure period i, observed at age j (O_{ij} can be cumulative paid losses, incremental paid losses, or incurred losses).
- C_{ij} — is known items (such as claim counts or inflation indices).
- F_j — is an index of reserve development factors, typically representing the percentage of the ultimate losses paid through j periods. This is estimated from the data.
- S_i — is an index reflecting the relative exposure at exposure period i. This is estimated from the data.
- K_{i+j} — is an index reflecting the relative effect of outside influences during a particular calendar period of time. This is estimated from the data.
- e_{ij} — are the differences between the observations O_{ij} and the estimated values of the process.

Since the C_{ij} are known items, it is possible to divide the O_{ij} by the C_{ij}. For example, C_{ij} may be the number of closed claims and O_{ij} the amount of paid losses. The parameter sets F_j, S_i and K_{i+j} will then effectively be estimating the average closed claim. Assuming that O_{ij} has been divided by C_{ij}, the parameter sets are chosen to model the observations as follows:

Exposure Period	Development Period			
	1	2	3	4
1	$F_1S_1K_1$	$F_2S_1K_2$	$F_3S_1K_3$	$F_4S_1K_4$
2	$F_1S_2K_2$	$F_2S_2K_3$	$F_3S_2K_4$	$F_4S_2K_5$
3	$F_1S_3K_3$	$F_2S_3K_4$	$F_3S_3K_5$	—
4	$F_1S_4K_4$	$F_2S_4K_5$	—	—
5	$F_1S_5K_5$	—	—	—

In a practical situation more than four development periods would be both available and desirable. The data which needs to be estimated to complete the reserve development is:

Exposure Period	Development Period			
	1	2	3	4
1	—	—	—	—
2	—	—	—	—
3	—	—	—	$F_4S_3K_6$
4	—	—	$F_3S_4K_6$	$F_4S_4K_7$
5	—	$F_2S_5K_6$	$F_3S_5K_7$	$F_4S_5K_8$

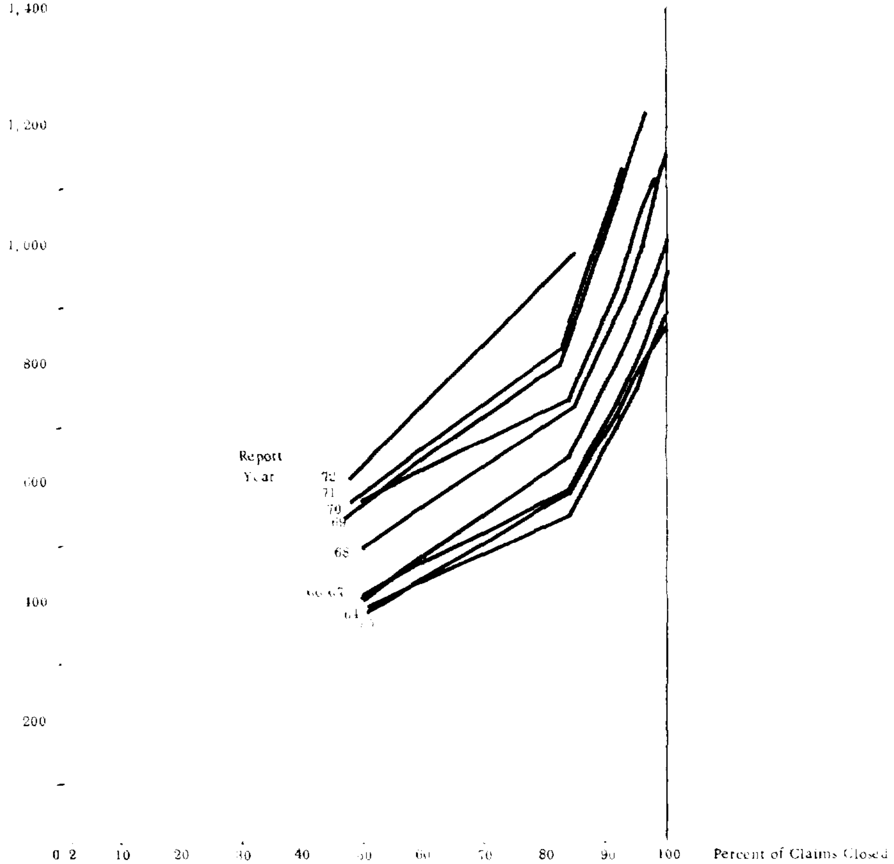
Additionally, some C_{ij} values might need to be estimated.

The F_j, S_i and K_{i+j} are parameter sets which are to be estimated subject to some criteria. They represent things unknown about the loss developments. The C_{ij} represent everything that is known or assumed to affect the developments. The C_{ij} can include measures of exposure, inflation, or claim counts. The C_{ij} could also represent changes in deductible levels or reinsurance retentions.

The F_j, S_i and K_{i+j} sets are all stated in terms of indices. Under certain circumstances they can be eliminated or replaced by functions. F_j has as many independent parameters as the number of development periods. S_i will have as many estimable parameters as the number of exposure periods. K_{i+j} will have as many estimable parameters as the larger of the number of

Figure 1.

CUMULATIVE AVERAGE CLOSED CLAIM



development periods or the number of exposure periods. In addition, to project the reserve developments several additional K_{i+j} terms must be projected.

There is an interesting interpretation of the various models which can be derived from the combination or elimination of F, S, and K parameter sets. Visualize the O_{ij} as incremental payments and ignore C_{ij}. The condition for F_j invariance then determines which parameters will be represented in the model. There are two choices of assumptions: (1) payments are in current or constant dollars, and (2) payments are related to the period of occurrence or the period of payment. Assume, for example, that F_j represents a constant percentage of payments in terms of constant dollars, valued at the occurrence of the claim. Observations will reflect the impact of inflation (current dollars) and a valuation at date of payment. The model thus needs a K_{i+j} term (to convert to constant dollars) and a S_i term (to value the claim at occurrence). The resulting model is thus $O_{ij} = C_{ij}F_jS_iK_{i+j} + e_{ij}$. Other variations are shown in the following table:

Value Of Money	Claim Valuation At	Model, $\hat{O}_{ij} =$
Current Dollars	Occurrence	F_jS_i
Current Dollars	Payment	F_j
Constant Dollars	Occurrence	$F_jS_iK_{i+j}$
Constant Dollars	Payment	F_jK_{i+j}

SOLUTION CRITERIA

There are many possible solution criteria to equation (1). For this paper the chosen criterion is to minimize the sum of squares of the differences between the observations and the estimates. The squares may be weighted by values a_{ij}, which might be chosen to reflect the relative credibility of the various observations. Algebraically the criterion is to minimize Z where:

$$Z = \sum_i \sum_j a_{ij}e_{ij}^2 = \sum_i \sum_j a_{ij}(O_{ij} - C_{ij}F_jS_iK_{i+j})^2.$$

The indices of summation apply to all available items.

A possible alternative, for example, might be Bailey and Simon's[4] minimum chi-square criterion. In the notation of this paper, that would be to minimize Z where:

$$Z = w \sum_i \sum_j a_{ij} \frac{(O_{ij} - C_{ij}F_jS_iK_{i+j})^2}{C_{ij}F_jS_iK_{i+j}}$$

where w is a constant. Bailey and Simon were concerned with rate equity, which can be reflected by the term in the denominator. In a sense, the minimum chi-square criterion attempts to minimize the error relative to the size of the observation. For loss reserving it is possible that the absolute error is more important than the relative error.

BASIC SOLUTION PROCEDURE

In order to derive a solution to the least squares formulation, it is generally necessary to make some simplifying assumptions. For this paper the following assumptions are made:

- The parameter sets F, S and K are independent of each other
- Individual index values within the parameter sets are independent of each other.

The O_{ij} and C_{ij} values are constants. The assumption of independence between different parameter sets is reasonable, since they are constructed to represent the three types of reserve developments (horizontal, vertical, and diagonal). As a practical matter, the S and K sets tend to be redundant. The independence of individual index values (particularly F and K) is in some doubt when the modelled data represents cumulative data.

[4] Bailey, R. A., and Simon, L. J., "Two Studies in Automobile Insurance Ratemaking," PCAS XLVII (1960).

In order to find the values of the parameters which will minimize the criterion function, set the partial derivative of the criterion function with respect to each parameter equal to zero. For the basic model and most variations the solution procedure will be iterative. To obtain a starting solution, one can assume that the F_j are the only parameters in the model.

The solution is thus:

$$\hat{F}_j = \frac{\sum_i a_{ij} O_{ij} C_{ij}}{\sum_i a_{ij} C_{ij}^2}$$

Next assume the model contains only F_j and S_i. Since F_j has been estimated above, it can be used to generate the initial estimates for S_i.

$$\hat{S}_i = \frac{\sum_j a_{ij} O_{ij} C_{ij} F_j}{\sum_j a_{ij} C_{ij}^2 F_j^2}$$

Finally, K_{i+j} can be solved:

$$\hat{K}_{i+j} = \frac{\sum_{i+j=c} a_{ij} O_{ij} C_{ij} F_j S_i}{\sum_{i+j=c} a_{ij} C_{ij}^2 F_j^2 S_i^2}$$

The revised values for F_j and S_i can then be found iteratively using analogous equations:

$$\hat{F}_j = \frac{\sum_i a_{ij} O_{ij} C_{ij} S_i K_{i+j}}{\sum_i a_{ij} C_{ij}^2 S_i^2 K_{i+j}^2}$$

$$\hat{S}_i = \frac{\sum_j a_{ij} O_{ij} C_{ij} F_j K_{i+j}}{\sum_j a_{ij} C_{ij}^2 F_j^2 K_{i+j}^2}$$

The computations proceed iteratively until no improvement in the criterion function can be made.

MODEL VARIATIONS

Some of the possible model variations include the choice of data. Observations can be (1) either paid or incurred losses, or (2) either incremental or cumulative developments. Further, C_{ij} can contain any variables known to affect the loss developments, including claim counts, premium, exposure, or inflation indices. Finally, various simplifications can be made for one of the parameter groups or they can be omitted.

A common assumption might be that inflation is a constant function of exposure period or of the calendar period. In these cases,

$$S_i = (1 + w)^i \text{ or}$$

$$K_{i+j} = (1 + w)^{i+j}$$

Fisher and Lange[5] assume that inflation will be constant for each age at settlement, or:

$$K_{i+j} = (1 + k_j)^i$$

For claim turnover intervals, a substitution is made for F_j, which can be of the general form (recognizing F_j can be bounded by 0 and 1):

$$F_j = 1 - \alpha + \alpha X_{ij}^{\beta}$$

where X_{ij} is the percentage of claims which have been closed.

SOLUTIONS TO MODEL VARIATIONS

The exponent introduced into the model variations makes it difficult to solve directly for the parameters. Newton's Method may be used. To find the minimum of the criterion function, one takes the derivative of the criterion with respect to each parameter and sets the resulting equation to zero. In other words:

$$f(k) = \frac{\partial Z}{\partial k} = 0$$

If k_1 is an initial estimate of k, a better estimate, k_2, can be found by Newton's Method as follows:

$$k_2 \cong k_1 - \frac{f(k_1)}{f'(k_1)}$$

5 *Ibid.*

The derivative can also be approximated as:

$$f'(k) \cong \frac{f(k+h) - f(k)}{h}$$

Thus: $$k_2 \cong k_1 - \frac{hf(k_1)}{f(k_1+h) - f(k_1)}$$

Initial parameter estimates can be obtained as described in an earlier section. The solution procedure will iterate while it successively estimates groups of parameters. When a parameter is estimated by Newton's Method, there will be a sub-iteration. Typical equations are given in the appendix.

NUMERICAL EXAMPLE

To compare the results of a variety of models, data from the Fisher-Lange paper are presented. Exhibit I shows the cumulative payments by reported year and development year; 84 months is considered the ultimate incurred loss. Also shown are the incremental average payments and the cumulative closed claim count. Complete data was not available in the original paper on the number of claims; it is therefore assumed that there are 1,000 claims per year.

Various models can be fitted to this data. For comparison, Exhibit II shows the estimated reserves for a variety of models. In each case the C matrix was taken as the number of closed claims. The K vectors were extrapolated based on a least-squares fit of the data points which could be estimated directly. Estimates are shown for both cumulative and incremental payments. The claim turnover approach can be shown graphically by Figure 1, where the cumulative average closed claim cost is shown as a function of the percentage of claims which have been closed. Equations for solving some of these models are given in the Appendix.

The example depicted in Exhibit II portrays some general results in the use of these models. First, models using K_{i+j} parameters are more difficult to use, since the parameters must be projected for future calendar year periods. In addition, closed claim counts must also be projected for future periods; with claim turnover intervals, however, closed claim count projections have no impact on the estimated reserve. Using cumulative data probably gives too much weight to early developments; thus, incremental data can lead to significantly different results. Estimating too many parameters yields arbitrary parameter values; for example, the inflation factors often add no explanatory value to that already provided by the S_i or K_{i+j} parameters; in addition, diagonal inflation can lead to the same result as exposure-period inflation.

Exhibit III shows the estimated parameters and projected developments for the F_jS_i model. Since the C_{ij} matrix is the cumulative closed claim count, the F_j vector is interpreted as the relative average claim value and the S_i vector is the average incurred claim cost for reported year i.

CONCLUSION

This paper presents a general approach to the modelling of loss reserve developments. All reserving methods are essentially mathematical models; all essentially assume that certain past events will be repeated in the future. This paper presents a methodology for understanding the assumptions made in any given model. In addition, it provides a means to generate a large number of alternative models. In particular, it stresses the use of all available information. This includes using the entire characteristic matrix format, instead of one observation or one diagonal; this also includes the use of endless types of collateral information, such as changes in deductible levels and external economic data. A general mathematical formulation is presented which allows the incorporation of all this data.

Most reserving methods are dependent upon certain fundamental assumptions, which may not be valid. How can one evaluate situations where: case reserving is inconsistent? the speed of claim settlements (payments) is changing? reinsurance retentions have changed? inflation is known to affect the data? Possible solutions to these questions will be briefly examined.

If case reserving is inconsistent, it may be best to evaluate only claim payments. If the rate of claim settlements are changing, the approach of claim turnover intervals is applicable. If reinsurance retentions have changed, adjustments can be incorporated into the C matrix. Inflation can be handled in a variety of ways. It can be assumed that inflation impacts either claim payments or claim occurrences. External economic functions or industrywide data can be used to model inflation. Finally, the inflation can be estimated from the claim data itself.

The value of the claim liability depends upon events which will occur in the future. A means of projecting the consequences of these events is to explore the various patterns which may continue with the future.

EXHIBIT I

INPUT DATA

I. Cumulative Payments

REPORTED YEAR	AGE 12	24	36	48	60	72	84
64	202184	465254	636658	726568	798577	829441	860385
65	197679	487722	660090	750090	807950	851498	891980
66	204848	489428	681620	795544	864044	899372	958682
67	224220	545194	760171	870281	930568	970908	1013188
68	247500	621480	823834	964447	1062748	1120069	1154607
69	286769	626641	852976	1026736	1153576	0	0
70	256695	658941	976191	1179090	0	0	0
71	275229	688579	1051977	0	0	0	0
72	291924	829946	0	0	0	0	0
73	350396	0	0	0	0	0	0

II. Incremental Average Payment

REPORTED YEAR	AGE 12	24	36	48	60	72	84
64	398	790	2348	2430	3429	2572	1934
65	393	871	2128	2500	2630	3629	3114
66	413	837	2288	2998	3425	2944	5931
67	444	961	2471	3146	3173	4034	4228
68	495	1084	2438	4261	4681	5211	4934
69	577	988	2865	4344	5285	0	0
70	545	1146	3375	4317	0	0	0
71	577	1181	3598	0	0	0	0
72	612	1466	0	0	0	0	0
73	698	0	0	0	0	0	0

EXHIBIT I (CONT'D)

III. Cumulative Closed Claim Count

REPORTED YEAR	AGE 12	24	36	48	60	72	84
64	508	841	914	951	972	984	1000
65	503	836	917	953	975	987	1000
66	496	836	920	958	978	990	1000
67	505	839	926	961	980	990	1000
68	500	845	928	961	982	993	1000
69	497	841	920	960	984	0	0
70	471	822	916	963	0	0	0
71	477	827	928	0	0	0	0
72	477	844	0	0	0	0	0
73	502	0	0	0	0	0	0

EXHIBIT II

COMPARATIVE SOLUTIONS FOR DIFFERENT MODELS (WHERE C_{ij} IS CLOSED CLAIM COUNT)

Model $O_{ij} = C_{ji}(\ \bullet\ )$	Number of Parameters	Estimated Reserve ($ Million) Cumulative Payments	Incremental Payments	Standard Error* (Cumulative)
F_jS_i	17	2.83	2.83	17.5
$F_j(1+w)^i$	9	2.77	2.81	25.6
$F_j(1+w)^{i+j}$	9	2.77	2.81	25.6
F_jK_{i+j}	17	2.83	2.83	17.8
$F_j(1+w_j)^i$	20	2.67	3.48	37.7
$X_{ij}(1-\alpha(1-X_{ij})^{\beta})S_i$	12	2.86		13.7

*Standard error calculated as square root of (sum of squares of differences between observations and projections divided by the number of observations less the number of parameters estimated) (shown only for cumulative payments model).

EXHIBIT III

MODEL OUTPUTS

I. Cumulative Payments

REPORTED YEAR	AGE 12	24	36	48	60	72	84
64	187943	453242	628325	732777	797809	829871	870945
65	191205	462923	647703	754488	822253	855266	894868
66	198851	488230	685346	799908	869871	904762	943787
67	218747	529400	745309	866965	941772	977548	1019713
68	244424	601730	842941	978420	1065012	1106561	1150804
69	256735	632844	883065	1032830	1127701	1165782	1216066
70	271106	689227	979693	1154446	1251454	1298994	1355024
71	284461	718426	1028321	1192352	1296584	1345838	1403889
72	317981	819590	1139583	1332852	1449366	1504425	1569315
73	350247	853733	1192710	1394989	1516935	1574560	1642476

II. Incremental Average Payment

REPORTED YEAR	AGE 12	24	36	48	60	72	84
64	370	797	2398	2823	3097	2672	2567
65	380	816	2281	2966	3080	2751	3046
66	401	851	2347	3015	3498	2908	3903
67	433	930	2482	3476	3937	3578	4216
68	489	1036	2906	4105	4123	3777	6320
69	517	1093	3167	3744	3953	6347	5028
70	576	1191	3090	3718	5706	4754	5603
71	596	1240	3068	5126	5212	4925	5805
72	667	1367	4210	4832	5826	5506	6489
73	698	1490	4237	5057	6097	5763	6792

III. Model Parameters

F_1	F_2	F_3	F_4	F_5	F_6	F_7
.425	.618	.789	.885	.942	.968	1.0

S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
871	895	944	1,020	1,151	1,216	1,355	1,404	1,569	1,642

APPENDIX

The following gives examples of models which can be solved by Newton's Method.

I. Model: $\hat{O}_{ij} = C_{ij}F_j(1 + w)^i$

Criterion: Minimize Z where $Z = \sum_i \sum_j a_{ij}(O_{ij} - \hat{O}_{ij})^2$

$$\frac{\partial Z}{\partial w} = 2 \sum_i \sum_j a_{ij}(O_{ij} - C_{ij}F_j(1 + w)^i)C_{ij}F_j(i - 1)(1 + w)^{i-1}$$

II. Model: $\hat{O}_{ij} = C_{ij}X_{ij}(1 - \alpha\,(1 - X_{ij}{}^{\beta}))S_i$

where X_{ij} is the cumulative fraction of exposure period i claims closed as of development period j.

Criterion: Minimize Z where $Z = \sum_i \sum_j a_{ij}(O_{ij} - O_{ij})^2$

$$\frac{\partial Z}{\partial \beta} = -2 \sum_i \sum_j a_{ij}[O_{ij} - C_{ij}S_iX_{ij}(1 - \alpha\,(1 - X^{\beta}_{ij}))]$$

$$C_{ij}S_iX_{ij}\,\alpha\,X_{ij}{}^{\beta}\ln X_{ij}$$