A Comparison of Stochastic Models that Reproduce Chain Ladder Reserve Estimates

Thomas Mack [a] and Gary Venter [b]

[a] *Munich Re, D-80791 München, email TMack@MunichRe.com*

[b] *Instrat, 120 Broadway, 29th Floor, New York, N.Y. 10271, email GVenter@SedgwickRe.BRKR.com*

Abstract

It is shown that the (over-dispersed) Poisson model is not the same as the distribution-free chain ladder model of Mack (1993) although both reproduce the historical chain ladder estimator for the claims reserve. For example, the true expected claims reserves, ignoring estimation issues, described by the two models are different. Moreover, the Poisson model deviates from the historical chain ladder algorithm in several aspects that the distribution-free chain ladder model does not. Therefore, only the latter can qualify to be referred to as the model underlying the chain ladder algorithm.

Keywords: chain ladder, distribution-free model, over-dispersed Poisson model

Overview of the two models

Let D_{ij} denote the cumulative losses for accident year i=1,...,n and for age j=1,...,n. The loss amounts D_{ij} have been observed for $j \leq n+1-i$ whereas the other amounts, especially the ultimate amounts D_{in}, i>1, have to be predicted. The chain ladder algorithm consists of the stepwise prediction rule

$$\hat{D}_{ij} = \hat{D}_{i,j-1}\,\hat{f}_j \quad \text{with} \quad \hat{f}_j = \sum_{i=1}^{n+1-j} D_{ij} \Big/ \sum_{i=1}^{n+1-j} D_{i,j-1}$$

starting with the most recent cumulative claims amount $\hat{D}_{i,n+1-i} = D_{i,n+1-i}$ of accident year i. This yields

$$\hat{D}_{in} = D_{i,n+1-i}\,\hat{f}_{n+2-i} \bullet \ldots \bullet \hat{f}_n$$

as estimator for the ultimate claims amount for year i.

The chain ladder algorithm was developed as a deterministic algorithm and did not have any stochastic model underlying it. Thus the question of what model underlies that algorithm cannot be answered by historical inquiry or even by strict logical deduction. But in order to assess its prediction error, an underlying stochastic model is required. In the last two decades, several models were associated with the chain ladder algorithm, but most of them lead to estimators for D_{in} which are different from the above $\hat{D}_{in}$, see e.g. Mack (1994) or England/Verrall (1998). The models compared in this paper are the only ones known that lead to the same estimators for D_{in} as the chain ladder algorithm. We argue that just one of them is close enough to the chain ladder algorithm in enough aspects that users of that algorithm should be able to accept all of its assumptions, and therefore should be comfortable with the prediction errors it implies. In this sense it qualifies to be called the stochastic model underlying the chain ladder algorithm.

Consider first the distribution-free stochastic model ("DFCL") of Mack (1993). Its main assumption is

(DFCL1) $\quad E(D_{ij} \mid D_{i1}, D_{i2}, \ldots, D_{i,j-1}) = D_{i,j-1} f_j \quad$ with unknown parameters $f_2,\ldots,f_n$.

From DFCL1 it can be deduced that

$$E(D_{in} \mid D_{i1},\ldots,D_{i,n+1-i}) = D_{i,n+1-i} f_{n+2-i} \bullet \ldots \bullet f_n$$

which immediately gives the chain ladder algorithm if the unknown parameters f_j are estimated by $\hat{f}_j$. It is important to realize that in claims reserving the relevant quantity to be estimated is the conditional mean of D_{in} given the data observed so far because these data are a part of the quantity D_{in} of interest. In Mack (1993) it is shown that $\hat{f}_j$ is the minimum variance unbiased linear estimator of f_j if one makes the following additional assumptions:

(DFCL2) $\quad Var(D_{ij} \mid D_{i1},\ldots,D_{i,j-1}) = D_{i,j-1}\sigma^2_j$ $\quad$ with unknown parameters σ^2_j.

(DFCL3) $\quad$ The accident years $(D_{i1},\ldots, D_{in})$, $1 \le i \le n$, are independent.

If DFCL2 is violated, some other variant of the chain ladder algorithm might still be optimal, as will be discussed further below. Failure of DFCL3 usually requires a different approach.

Already in 1975, Hachemeister and Stanard discovered another stochastic model leading to the chain ladder algorithm. This model works on the incremental amounts

$$C_{i1} = D_{i1} \quad \text{and} \quad C_{ij} = D_{ij} - D_{i,j-1}\,,\; j > 1,$$

and makes the assumptions

(P1) $\quad E(C_{ij}) = x_i y_j$ with unknown parameters x_i and y_j.

(P2) $\quad$ Each C_{ij} has a Poisson distribution.

(P3) $\quad$ All C_{ij} are independent.

Then the maximum likelihood estimator $\hat{R}_i = \hat{x}_i\hat{y}_{n+2-i} + \ldots + \hat{x}_i\hat{y}_n$ of the claims reserve for year i,

$$R_i = C_{i,n+2-i} + \ldots + C_{in} = D_{in} - D_{i,n+1-i}$$

turns out to give the same prediction $\hat{D}_{in} = D_{i,n+1-i} + \hat{R}_i$ as the chain ladder algorithm. Due to the independence assumption P3, $D_{i,n+1-i} + \hat{R}_i$ is an estimator of the conditional expectation $E(D_{in} \mid D_{i1},\ldots,D_{i,n+1-i})$. Assumption P2 implies that all incremental amounts C_{ij} have to be non-negative integers.

This Poisson model can be cast into the form of a Generalized Linear Model (GLM) with logarithmic link function via (cf. Renshaw/Verrall (1994))

(P1*) $\quad \ln(E(C_{ij})) = \alpha_i + \beta_j$.

In this form, the parameters $\alpha_i = \ln(x_i)$, $\beta_j = \ln(y_j)$ can be estimated using standard statistical GLM software. Of course, this yields the same estimators for x_i, y_j, R_i and D_{in} because the estimation procedure of GLMs is maximum likelihood, too. A further benefit of the use of GLMs is the fact that these show a way to overcome the constraint that C_{ij} have to be non-negative integers. For this purpose, we recall that the only distributional assumptions used in GLMs are the functional relationship between variance and mean and the fact that the distribution belongs to the exponential family. This relationship is $Var(C_{ij}) = E(C_{ij})$ in the Poisson case. It can be generalized to $Var(C_{ij}) = \phi E(C_{ij})$ without any change in form and solution of the likelihood equations. This allows for more dispersion in the data. For the solution of the likelihood equations it is not necessary that the C_{ij}'s are non-negative or integers. Therefore one does not care any more about the range of the underlying distribution; one speaks of an over-dispersed Poisson model and of quasi-likelihood equations. At least, the fitted values are always positive. However, the algorithm for the solution of the quasi-likelihood equations breaks down unless the sums of the observed incremental amounts in every row and every column are non-negative as can easily be seen from the quasi-likelihood equations

$$\sum_{j=1}^{n+1-i} \exp(\alpha_i + \beta_j) = \sum_{j=1}^{n+1-i} C_{ij}, \qquad 1 \le i \le n,$$

$$\sum_{i=1}^{n+1-j} \exp(\alpha_i + \beta_j) = \sum_{i=1}^{n+1-j} C_{ij}, \qquad 1 \le j \le n.$$

But this problem can be overcome if we work without the log-link and without GLM software. Then the over-dispersed Poisson model ("ODP") is:

(ODP1) $E(C_{ij})) = x_i y_j$ with unknown parameters x_i, y_j.

(ODP2) The distribution of C_{ij} belongs to the exponential family with $Var(C_{ij}) = \phi E(C_{ij})$ where ϕ is an unknown parameter.

(ODP3) All C_{ij} are independent.

The resulting quasi-likelihood equations are

$$\sum_{j=1}^{n+1-i} x_i y_j = \sum_{j=1}^{n+1-i} C_{ij}, \qquad 1 \le i \le n,$$

$$\sum_{i=1}^{n+1-j} x_i y_j = \sum_{i=1}^{n+1-j} C_{ij}, \qquad 1 \le j \le n.$$

As has been shown in Appendix A of Mack (1991) (see also Schmidt/Wünsche (1998)), these equations have the unique solution (if all $\hat{f}_j$ are well-defined and $\neq 0$, but without any restrictions on the row sums or column sums over C_{ij})

$$\hat{x}_i \hat{y}_j = D_{i,n+1-i} \hat{f}_{n+2-i} \cdot \ldots \cdot \hat{f}_{j-1} \cdot (\hat{f}_j - 1) \qquad \text{for } j > n+1-i,$$

$$\hat{x}_i \hat{y}_j = D_{i,n+1-i}((\hat{f}_{j+1} \cdot \ldots \cdot \hat{f}_{n+1-i})^{-1} - (\hat{f}_j \cdot \ldots \cdot \hat{f}_{n+1-i})^{-1}) \qquad \text{for } j \le n+1-i,$$

with $\hat{f}_j$ from the chain ladder algorithm. Because $(\hat{f}_j - 1) + \hat{f}_j \cdot (\hat{f}_{j+1} - 1) = \hat{f}_j \cdot \hat{f}_{j+1} - 1$ we immediately obtain

$$\hat{R}_i = \hat{x}_i \hat{y}_{n+2-i} + \ldots + \hat{x}_i \hat{y}_n = D_{i,n+1-i}(\hat{f}_{n+2-i} \cdot \ldots \cdot \hat{f}_n - 1).$$

This shows that the solution of the quasi-likelihood equations of the ODP model gives the same estimator for D_{in} as the chain ladder algorithm and as the DFCL model. We will find some restrictions later.

Evidence that the models are different

In view of the fact that both models, DFCL1-3 and ODP1-3, yield the same estimators $\hat{R}_i$ and $\hat{D}_{in}$ as the deterministic chain ladder algorithm, the questions arise whether one model is a special case of the other and whether both models can be called "underlying the chain ladder algorithm". First, we give five arguments, each of which shows that the models are different.

a) ODP has more parameters than DFCL:

In ODP1, one of the parameters is redundant because replacing x_i with x_i/c and y_j with $y_j \cdot c$ yields the same model. Therefore ODP has 2n-1 parameters whereas DFCL has only n-1 parameters $f_2, \ldots, f_n$. If we make the parametrisation of ODP1 unique by requiring that $y_1 = 1$, then we have a one-to-one relationship between the set of the $\hat{y}_j$ and the set of the $\hat{f}_j$:

$$\hat{f}_j = (\hat{y}_1 + \ldots + \hat{y}_j) / (\hat{y}_1 + \ldots + \hat{y}_{j-1})$$

$$\hat{y}_j = \hat{f}_2 \cdot \ldots \cdot \hat{f}_{j-1} \cdot (\hat{f}_j - 1), \quad j > 1.$$

The parameters x_i of ODP have an equivalent neither in the chain ladder algorithm nor in the DFCL model. Within ODP, they make it possible to estimate the unconditioned means $E(D_{ij})$, too, whereas within the DFCL model only the conditional means $E(D_{ij} \mid D_{i1}, \ldots, D_{i,j-1})$ can be estimated.

Sometimes it is claimed, each $D_{i,n+1-i}$ would be a further parameter in the chain ladder algorithm, being an estimator for $E(D_{i,n+1-i})$. But this argument is not convincing because there would exist better estimators, e.g. any average of all $D_{ij}\hat{f}_{j+1}\bullet\ldots\bullet\hat{f}_{n+1-i}$, $1 \leq j \leq n+1-i$. Moreover, the DFCL model clearly shows that there $D_{i,n+1-i}$ is not an estimator of $E(D_{i,n+1-i})$ but is the result of conditioning on the known data.

b) ODP and DFCL have diverging independence assumptions:

Either model assumes that the accident years (rows) are independent. In addition, ODP3 requires all increments C_{ij} within each accident year to be independent, too, whereas DFCL implies that all C_{ij} within the same accident year are correlated: e.g. given $D_{i,j-1}$, we have (with $E(\varepsilon_{j+1}) = 0$)

$$C_{i,j+1} = D_{i,j+1} - D_{ij} = D_{ij}(f_{j+1}-1) + \varepsilon_{j+1} = (D_{i,j-1} + C_{ij})(f_{j+1}-1) + \varepsilon_{j+1}$$

which shows that $C_{i,j+1}$ and C_{ij} are correlated.

c) The fitted values $\hat{C}_{ij}$ or $\hat{D}_{ij}$ and therefore also the residuals $r_{ij} = \hat{C}_{ij} - C_{ij}$ for ODP are different from those of DFCL:

The fitted value $\hat{D}_{ij}$ for $j \leq n+1-i$ for the DFCL model clearly is

$$\hat{D}_{ij} = D_{i,j-1}\hat{f}_j \quad \text{or} \quad \hat{C}_{ij} = D_{i,j-1}(\hat{f}_j - 1).$$

Note that there are no fitted values for the first column $D_{i1} = C_{i1}$, $1 \leq i \leq n$.

The fitted value $\hat{D}_{ij}$ for $j \leq n+1-i$ for the ODP model is (see above)

$$\hat{C}_{ij} = \hat{x}_i\hat{y}_j = D_{i,n+1-i}\left((\hat{f}_{j+1}\bullet\ldots\bullet\hat{f}_{n+1-i})^{-1} - (\hat{f}_j\bullet\ldots\bullet\hat{f}_{n+1-i})^{-1} \right),$$

which yields

$$\hat{D}_{ij} = D_{i,n+1-i} / (\hat{f}_{j+1}\bullet\ldots\bullet\hat{f}_{n+1-i})$$

which is different from DFCL, because e.g. for j=n+1-i we obtain $D_{i,n-i} \cdot \hat{f}_{n+1-i}$ as fitted value for DFCL and $D_{i,n+1-i}$ as fitted value for ODP which are different for i > 1 because $\hat{f}_{n+1-i}$ depends not only on $D_{i,n+1-i}/D_{i,n-i}$ of accident year i but also on the corresponding ratios of all older accident years, too. Another way to see the difference is to look at $\hat{D}_{ij} / \hat{D}_{i,j-1}$ for j fixed which is constant for ODP ($=\hat{f}_j$) but not for DFCL.

d) The true expected reserves $E(R_i \mid data)$ described by the models are different:

If we would know the true parameters $f_2, \ldots, f_n$ and x_i, y_j, respectively and if it were possible to draw several sets of observations (run-off triangles $\{D_{ij},\ i+j \leq n+1\}$) from the same "population", then the reserve estimator (= expected reserve) $\hat{R}_i = E(R_i \mid data) = D_{i,n+1-i}(f_{n+2-i} \cdot \ldots \cdot f_n - 1)$ of the chain ladder algorithm as well as of the DFCL model would change from one random draw to the next as $D_{i,n+1-i}$ changes, but the reserve estimator $\hat{R}_i = E(R_i \mid data) = x_i y_{n+2-i} + \ldots + x_i y_n$ of the ODP model would not.

e) The simulated future emergence is different:

To simulate the future values of C_{ij} for ODP, all their expected values could be calculated, and then random draws added. For DFCL, on the other hand, the first new diagonal could be simulated this way, but for the second new diagonal the simulated cumulative value for the first diagonal would have to be multiplied by the development factor to get the mean value for the second diagonal. This mean thus includes the random component simulated for the first diagonal, which is not the case for the ODP simulation.

Conclusion:

Whereas after a) ODP could still seem to be a special case of DFCL, b), c), d) and e) show that the models are definitively different. In each of these areas of difference, the DFCL agrees with the chain ladder algorithm to the extent that the algorithm has a clear implication. For instance, the chain ladder algorithm does not compute any parameters other than the development factors, and if you wanted to measure goodness-of-fit for the chain ladder algorithm, you would calculate the fitted values as the development factors times the previous cumulative losses. The ODP differs from the chain ladder algorithm in these aspects.

Moreover, the fact that the true expected reserves are different, clearly shows that the reserve estimators of both models coincide only incidentially and not due to any inner connection. As a consequence, especially from d) and e), the prediction errors of the models are different, too. For the formulae of the prediction errors see Mack (1993) for the DFCL model and England/Verrall (1998) for the ODP model. Therefore, the prediction errors calculated with the bootstrap approach of England and Verrall (1998) cannot be viewed to be the prediction errors for the chain ladder algorithm. A bootstrap approach to the chain ladder predicton errors would contain a resampling the individual development factors $F_{ij} = D_{ij}/D_{i,j-1}$ within each column j, but would be questionable for the last few columns, due to limited data.

Other differences between the ODP model and the chain ladder algorithm

The chain ladder algorithm leads to closely related alternatives when the data is not exactly as required, and these alternatives can be readily accommodated by the DFCL model with appropriate adjustments to the assumptions. This is not the case with the ODP in the following two examples.

1) ODP does not work in all situations where the chain ladder algorithm works:

In practise, it often happens that the figures of the oldest calendar years are not available. This means that the run-off triangle $\{D_{ij} \mid i+j \leq n+1\}$ changes its shape to a trapezoid $\{D_{ij} \mid m \leq i+j \leq n+1\}$ with $2 < m \leq n$. Then the ODP model does not yield estimators $\hat{R}_i$ identical to the DFCL model and chain ladder algorithm. This is most easily seen for the following trapezoid (missing D_{11}):

	D_{12}	D_{13}	with incremental amounts			C_{13}
D_{21}	D_{22}			C_{21}	C_{22}	
D_{31}				C_{31}		

where C_{12} cannot be known because otherwise one would also know $D_{11} = D_{12} - C_{12}$. In this situation the ODP model clearly cannot be applied (4 observations for 5 parameters) but both, chain ladder algorithm and DFCL model, can (with $\hat{f}_2 = D_{22}/D_{21}$, $\hat{f}_3 = D_{13}/D_{12}$) and yield identical results. Here, the disadvantage of the additional parameters x_i of the ODP is clearly exhibited.

More generally, if one or several claims amounts D_{ij} or C_{ij} somewhere in the run-off triangle are not known or omitted as being not credible (e.g. a typing error or a large individual claim whose amount is not known), then the ODP yields different estimations for D_{in} from chain ladder algorithm (and from DFCL). The reason is the fact that the proof for the formula

$$\hat{x}_i \hat{y}_j = D_{i,n+1-i} \cdot \hat{f}_{n+2-i} \cdot \ldots \cdot \hat{f}_{j-1} \cdot (\hat{f}_j - 1) \qquad \text{for } j > n+1-i,$$

only works if $D_{i,n+1-i}$ equals the sum of all C_{ij} known for accident year i which is not the case if some C_{ij} are missing.

2) ODP can not be adapted if the weights in $\hat{f}_j$ are changed:

In practise, the estimator for f_j is often calculated in a different way from $\hat{f}_j$, e.g.

$$'\hat{f}_j = \frac{1}{n+1-j} \sum_{i=1}^{n+1-j} \frac{D_{ij}}{D_{i,j-1}} \qquad \text{(straight average)}$$

or

$$''\hat{f}_j = \sum_{i=1}^{n+1-j} D_{i,j-1} D_{ij} \Big/ \sum_{i=1}^{n+1-j} D_{i,j-1}^2 \qquad \text{(linear regression, cf. Kremer (1984)).}$$

These estimators lead to different estimators for R_i and D_{in} from those based on $\hat{f}_j$. It can be shown that the DFCL model can easily be adapted to these estimators via

(DFCL2') $\quad Var(D_{ij} \mid D_{i1},\ldots,D_{i,j-1}) = D_{i,j-1}^2 \, \sigma_j^2$

or

(DFCL2") $\quad Var(D_{ij} \mid D_{i1},\ldots,D_{i,j-1}) = \sigma_j^2$.

But if one tries to adapt ODP3 to

(ODP3') $\quad Var(C_{ij}) = \phi(E(C_{ij}))^2$ $\qquad$ (Gamma-GLM, cf. Mack (1991))

or

(ODP3") $\quad Var(C_{ij}) = \phi$ $\qquad$ (least squares method of DeVylder (1978)),

one obtaines estimators for R_i and D_{in} which are again different from those of the CL algorithm with $'\hat{f}_j$ or $''\hat{f}_j$, respectively. There are no other variance assumptions known under which ODP could be adapted to give the same predictions as the chain ladder algorithm with $'\hat{f}_j$ or $''\hat{f}_j$.

Conclusion

The DFCL and ODP models are different and they yield the same predictions only in a special but common situation. But even then, the prediction errors and the model descriptions of the true amounts to be predicted are different. Because the DFCL model agrees with the chain ladder algorithm in every identified area, whereas the ODP does not, users of that algorithm should be able to accept all of the assumptions of the DFCL. In fact, Venter (1998) shows that when those assumptions are violated, other algorithms should be considered. Thus when a stochastic model is needed to estimate chain ladder prediction errors, only the DFCL model can considered to be underlying the chain ladder algorithm.

References

DeVylder (1978), Estimation of IBNR Claims by Least Squares, Mitteilungen der SVVM 78, 249-254.

England and Verrall (1998), Standard Errrors of Prediction in Claims Reserving: A Comparison of Methods, Proceedings of the General Insurance Convention & ASTIN Colloquium in Glasgow, Volume 1, 459-478

Hachemeister and Stanard (1975), IBNR Claims Count Estimation with Static Lag Functions, Spring Meeting of the Casualty Actuarial Society.

Kremer (1984), A Class of Autoregressive Models for Predicting the Final Claims Amount, IME 3, 111-119.

Mack (1991), A Simple Parametric Model for Rating Automobile Insurance or Estimating IBNR Claims Reserves, ASTIN Bulletin 21, 93-109.

Mack (1993), Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates, ASTIN Bulletin 23, 213-225.

Mack (1994), Which Stochastic Model is Underlying the Chain Ladder Method?, IME 15, 133-138.

Renshaw and Verrall (1994), A Stochastic Model Underlying the Chain-Ladder Technique, Proceedings of the ASTIN Colloquium in Cannes, 45-69

Schmidt and Wünsche (1998), Chain Ladder, Marginal Sum and Maximum Likelihood Estimation, Blätter der DGVM 23, 267-277.

Venter (1998), Testing the Assumptions of Age-to-Age Factors, Proceedings of the Casualty Actuarial Society LXXXV.