CHAIN LADDER BIAS

BY

GREG TAYLOR

ABSTRACT

The chain ladder forecast of outstanding losses is known to be unbiased under suitable assumptions. According to these assumptions, claim payments in any cell of a payment triangle are dependent on those from preceding development years of the same accident year. If all cells are assumed stochastically independent, the forecast is no longer unbiased. Section 5 shows that, under rather general assumptions, it is biased upward. This result is linked to earlier work on some stochastic versions of the chain ladder.

KEYWORDS

Chain ladder, IBNR.

1. INTRODUCTION

The chain ladder (CL) approach to estimation of a loss reserve is well known. It is described, for example, by Taylor (2000).

Its origins are not altogether clear, but it seems likely that it originated as a heuristic device. As such, it may be viewed as a non-parametric estimator. The precise definition is given in Sections 2 and 3.

Kremer (1982) recognised that the CL involved a log-linear cross-classification structure. A number of parametric stochastic versions of the CL developed from this, eg Hertig (1985), Renshaw (1989), Verrall (1989, 1990, 1991).

Mack (1994) pointed out that these stochastic models gave mean estimates of liability that differed from the "classical" CL estimate. While the form of stochastic model underlying the classical CL was speculative, due to the latter's heuristic nature, Mack suggested one. It is distribution free. Details are given in Section 2. Mack also identified the differences between this and the other stochastic models.

Whereas the cross-classified models typically assume stochastic independence of all cells in the data set, the CL (in Mack's formulation) does not. It was shown by Mack (1993) that the algorithm of the classical CL produced unbiased forecasts of liability under its own assumptions.

However, it does not necessarily do so under the alternative assumption of independence between all cells. Some papers have studied the bias in estimates

https://doi.org/10.2143/AST.33.2.503695 Published online by Cambridge University Press

of liability in the parametric cross-classified models mentioned above, but little is known of the bias in the classical CL forecast when all cells are independent.

The purpose of the present paper is to investigate the direction of bias in this case.

2. Framework and notation

Consider a square array X of stochastic quantities $X(i,j) \geq 0$, $i = 0,1,\ldots,I$; $j = 0,1,\ldots,I$.

Denote row sums and column sums as follows:

$$R(i,j) = \sum_{h=0}^{j} X(i,h) \tag{2.1}$$

$$C(i,j) = \sum_{g=0}^{i} X(g,j). \tag{2.2}$$

In addition introduce the following notation for the total sum over a rectangular subset of X:

$$\begin{aligned} T(i,j) &= \sum_{g=0}^{i} \sum_{h=0}^{j} X(g,h) \\ &= \sum_{g=0}^{i} R(g,j) \\ &= \sum_{h=0}^{j} C(i,h) \end{aligned} \tag{2.3}$$

Generally, in the following, any summation of the form $\sum_{a}^{b}$ with $b < a$ will be taken to be zero.

In a typical loss reserving framework, i denotes accident period, j development period, and available data will consist of observations on the triangular subset Δ of X:

$$\Delta = \{X(i,j), i = 0,1,\ldots,I;\ j = 0,1,\ldots,I-i\} \tag{2.4}$$

Figure 2.1 illustrates the situation.

Still in a loss reserving context, Δ would represent some form of claims experience, eg claim counts or claim amounts. The loss reserving problem consists of forecasting the lower triangle in Figure 2.1, conditional on Δ. There is particular interest in forecasting $R(i,I) \mid \Delta$, $i = 1,\ldots,I$. In standard loss reserving parlance, the $X(i,j)$ are usually referred to as **incremental** quantities, or just **increments**, and the $R(i,j)$ as their **cumulative** equivalents.

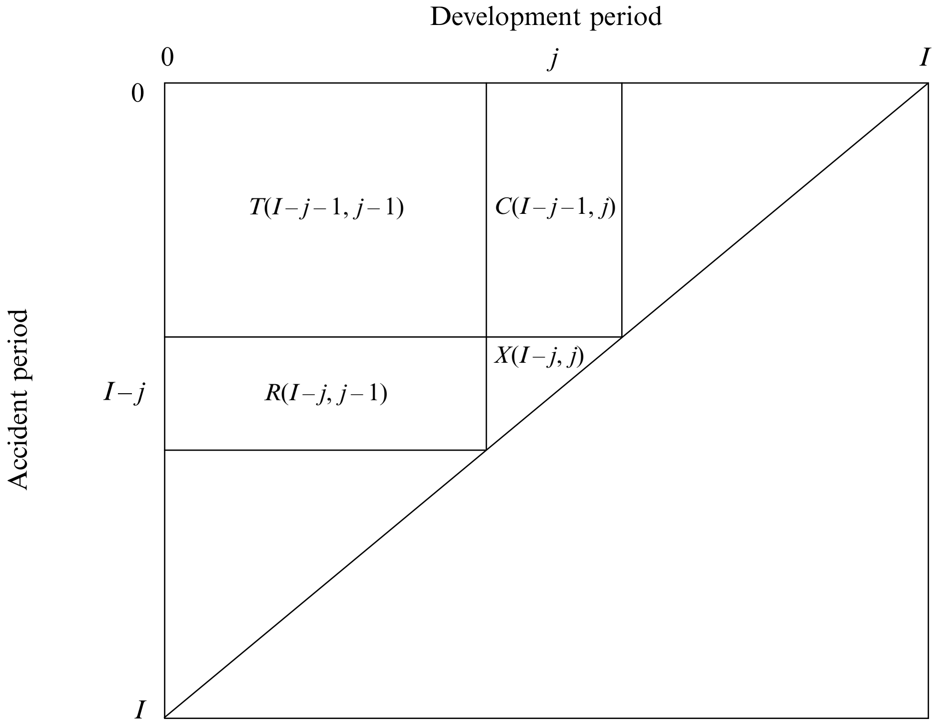

Figure 2.1: Data array

Matrix notation

Analysis of CL bias will proceed by examination of certain derivatives of the CLF defined in Section 3. These depend ultimately on derivatives of the $\hat{v}(k)$. Quantities such as $\partial^2\hat{v}(k)/\partial^2 X(i,j)$ require evaluation, and a matrix notation will be useful for keeping organization among the indexes.

In the following, all vectors will be of dimension $(I+1)$, and all matrices of dimension $(I+1)\times(I+1)$.

Let X denote the matrix $[X(i,j)], i,j = 0,1,\ldots,I$. Also let e_k denote the natural basis vector:

$$e_k^T = \Big(0,\ldots,0,\underset{(k)}{1},0,\ldots,0\Big),\ k = 0,1,\ldots,I \tag{2.5}$$

and define

$$f_k^T = \sum_{r=0}^{k} e_k^T = \Big(1,\ldots,\underset{(k)}{1},0,\ldots,0\Big) \tag{2.6}$$

where the upper T denotes transposition.

Under this notation, $X(i,j)$ is selected from X as follows:

$$X(i,j) = e_i^T X e_j. \tag{2.7}$$

Similarly,

$$R(i,j) = e_i^T X f_j \tag{2.8}$$

$$C(i,j) = f_i^T X e_j \tag{2.9}$$

$$T(i,j) = f_i^T X f_j. \tag{2.10}$$

3. Chain ladder forecast

Define the **age-to-age factor**

$$\begin{aligned}\hat{\nu}(j) &= T(I-j-1, j+1)/T(I-j-1, j) \\ &= 1 + C(I-j-1, j+1)/T(I-j-1, j)\end{aligned} \tag{3.1}$$

and

$$\hat{R}(i,I) = R(i,I-i)\prod_{k=I-i}^{I-1}\hat{\nu}(k). \tag{3.2}$$

The value of $\hat{R}(i,I)$ calculated in this way will be referred to as the **chain ladder forecast** (CLF) of $R(i,I)$.

4. Chain ladder models

4.1. Independent accident periods

The CLF has been formulated in Section 3 just as an algorithm. No model for the data X has yet been stated.

It is evident that the properties of the CLF will depend on the model. This and the next sub-section consider two alternative models. The first is represented by the following two assumptions.

Assumption 1. The increments of different accident periods are stochastically independent in the sense that

$$Prob[X(i,j) \mid \Delta] = Prob[X(i,j) \mid X(i,0), \ldots, X(i,I-i)].$$

Assumption 2. $E[R(i,j+1) \mid X(i,0), X(i,1), \ldots, X(i,j)] = \nu(j)R(i,j).$ (4.1)

https://doi.org/10.2143/AST.33.2.503695 Published online by Cambridge University Press

Remark 1. It follows from Assumptions 1 and 2 that

$$E[R(i,j+1)\,|\,\Delta] = \nu(j)R(i,j) \tag{4.2}$$

for any $j \geq I-i$ (ie future j).

Since $R(i,j+1) = R(i,j) + X(i,j+1)$, one may re-write (4.2) in the form:

$$E[X(i,j+1)\,|\,\Delta] = [\nu(j)-1]R(i,j) \tag{4.3}$$

Remark 2. It is clear from (4.3) that $R(i,j)$ and $X(i,j+1)$ may fail to be independent. Generally, under Assumptions 1 and 2, the $X(i,j)$ for fixed i may fail to be independent.

Theorem 1 (Mack). Under Assumptions 1 and 2,

(1) $\hat{\nu}(j)$ is an unbiased estimator of $\nu(j)$ for $j = 0,1,\ldots,I-1$; and
(2) the CLF $\hat{R}(i,I)$ is an unbiased estimator of $E[R(i,I)\,|\,\Delta]$ for $i = 1,2,\ldots,I$;

provided that both estimators exist.

Proof. See Mack (1993). □

It is also convenient to re-write (4.2) in the form:

$$E\big[R(i,j+1)/R(i,j)\,|\,\Delta\big] = \nu(j). \tag{4.4}$$

4.2. Independent increments

Replace Assumptions 1 and 2 by 1a, 2a and 3 as follows.

Assumption 1a. All increments $X(i,j)$ are stochastically independent.

Assumption 2a. $E[R(i,j+1)]/E[R(i,j)] = \eta(j)$. (4.5)

The assumption is written in this form in order to relate it to Assumption 2. It is useful to note that an equivalent, and more natural, assumption is that

$$E[X(i,j)] = \alpha(i)\beta(j) \tag{4.5a}$$

for parameters $\alpha(i), \beta(j)$, $i,j = 0,1,\ldots,I$.

Define the set

$$D_i = \{(g,h): g \leq I-k-1, h \leq k+1,\; k = I-i,\ldots,I-1\} \tag{4.6}$$

and

$$E_i = \{(g,h): g \leq I-k-1, h \leq k,\; k = I-i,\ldots,I-1\}.$$

https://doi.org/10.2143/AST.33.2.503695 Published online by Cambridge University Press

Assumption 3. $T(g,h) > 0$ for each $(g,h) \in E_i$.

Remark 3. It is implicit in Assumption 2a that $E[R(i,j)] \neq 0$. By the assumed non-negativity of the $X(i,j)$, $E[R(i,j)] > 0$ for each i, j.

Remark 4. A comparison of (4.4) and (4.5) indicates that $\nu(j)$ and $\eta(j)$ are different quantities (for fixed j) since

$$E\left[R(i,j+1)/R(i,j)\right] \neq E\left[R(i,j+1)\right]/E\left[R(i,j)\right]. \tag{4.7}$$

This fact was pointed out by Mack (1994).

By Assumption 3, applied to (3.1), all $\hat{\nu}(k)$ appearing in (3.2) are defined and strictly positive. Given the assumed non-negativity of the $X(i,j)$, Assumption 3 is both necessary and sufficient for the chain ladder forecast to make sense. Comment on the non-negativity requirement will be made in Section 6.

The conditions of Theorem 1 no longer hold, and so the CLF is not necessarily unbiased.

5. Chain ladder bias

The following is a somewhat technical result, but has been included here rather than in the appendix because the symmetric appearance of rows and columns of X in the second derivative of $Y(i)$ is interesting.

Theorem 2. Define

$$Y(i) = \prod_{k=I-i}^{I-1} \hat{\nu}(k). \tag{5.1}$$

Then

$$\frac{\partial^2 Y(i)}{\partial X^2(g,h)} = 0 \text{ for } (g,h) \notin D_i; \tag{5.2}$$

for $(g,h) \in D_i$ and $h \leq I-i$,

$$\frac{1}{Y(i)} \frac{\partial^2 Y(i)}{\partial X^2(g,h)} = 2 \sum_{k=I-i}^{I-g-1} \frac{C(I-k-1,k+1)}{T(I-k-1,k+1)\,T(I-k-1,k)} \text{ x } \left[\frac{1}{T(i-1,I-i)} + \sum_{l=I-i+1}^{k} \frac{R(I-l,l)}{T(I-l-1,l)\,T(I-l,l)}\right] \tag{5.3}$$

https://doi.org/10.2143/AST.33.2.503695 Published online by Cambridge University Press

for $(g,h) \in D_i$ and $h > I-i$,

$$\frac{1}{Y(i)}\frac{\partial^2 Y(i)}{\partial X^2(g,h)} = 2\sum_{k=h}^{I-g-1}\frac{C(I-k-1,k+1)}{T(I-k-1,k+1)\,T(I-k-1,k)} \times \sum_{l=h}^{k}\frac{R(I-l,l)}{T(I-l-1,l)\,T(I-l,l)}. \tag{5.4}$$

These results depend on Assumption 3 for the existence of (5.3) and (5.4), but do **not** depend on the Assumptions 1a and 2a.

Proof. See appendix. □

Theorem 3. Under Assumptions 1a, 2a and 3, and if $X(g,h)$ is not degenerate for at least one $(g,h) \in D_i$, the CLF $\hat{R}(i,I)$ is biased upward as an estimate of $E[R(i,I)]$ in the sense that

$$E\left[\hat{R}(i,I)\middle|R(i,I-i)\right] > R(i,I-i)\frac{E[R(i,I)]}{E[R(i,I-i)]}. \tag{5.5}$$

Also,

$$E\left[\hat{R}(i,I)\right] > E[R(i,I)]. \tag{5.6}$$

Proof. See appendix. □

Remark 5. An alternative form of (5.6) is:

$$E_{R(i,I-i)}E\left[\hat{R}(i,I)\middle|R(i,I-i)\right] > E[R(i,I)]. \tag{5.7}$$

Note that, in general, (5.5) does **not** imply that

$$E\left[\hat{R}(i,I)\middle|R(i,I-i)\right] > E[R(i,I)].$$

A few words in interpretation of Theorem 3. First define

$$\begin{aligned} L(i,j) &= R(i,I) - R(i,j) \\ &= \sum_{h=j+1}^{I} X(i,h) \end{aligned} \tag{5.8}$$

which is the required loss reserve in respect of accident period i at the end of development period j.

The forecast of $L(i,j)$ associated with the CLF of $R(i,I)$ is

$$\hat{L}(i,j) = \hat{R}(i,I) - R(i,j). \tag{5.9}$$

Now consider the CLF $\hat{L}(i, I-i)$, taken on the last diagonal of Δ, and how it is conditioned by Δ. By (3.2),

$$\hat{L}(i,I-i) = R(i,I-i)\left[\prod_{k=I-i}^{I-1}\hat{\nu}(k)-1\right] \tag{5.10}$$

ie is proportional to $R(i,I-i)$.

By Theorem 1,

$$E\left[\hat{L}(I-i)\,\middle|\,\Delta\right] = E\left[L(I-i)\,|\,\Delta\right]. \tag{5.11}$$

in the **dependent case**, when Assumptions 1 and 2 hold. The CLF of loss reserve is **conditionally unbiased**.

On the other hand, in the **independent case**, when Assumption 1a holds, $L(i,j)$ and $R(i,j)$ are stochastically independent since they involve disjoint sets of the $X(i,h)$ (see (2.1) and (5.8)). Hence

$$E\left[L(i,I-i)\,|\,\Delta\right] = E\left[L(i,I-i)\right]. \tag{5.12}$$

If Assumptions 2a and 3 also hold, then Theorem 3 may be applied to (5.12) to yield

$$E\left[\hat{L}(i,I-i)\right] > E\left[L(i,I-i)\,|\,\Delta\right] = E\left[L(i,I-i)\right] \tag{5.13}$$

In this case, the CLF of loss reserve, taken **unconditionally**, is **biased upward**. Note that it makes no sense to discuss whether the CLF is conditionally biased in this case. For (5.10) shows that

$$E\left[\hat{L}(i,I-i)\,\middle|\,\Delta\right] \text{ is proportional to } R(i,I-i) \tag{5.14}$$

whereas (5.12) shows that

$$E\left[L(i,I-i)\,|\,\Delta\right] \text{ is independent of } R(i,I-i). \tag{5.15}$$

Then, whether $\hat{L}(i,I-i)$ is conditionally biased upward or downward depends on $R(i,I-i)$.

6. Negative increments

It has been assumed throughout that all $X(i,j) \geq 0$. It is evident that some positivity assumption is required for Theorem 3 to hold. If, for example, one were to apply the theorem to data that were subject to Assumption 3, and then reverse the signs of all $X(i,j)$, one would obtain a **downward** bias for the CLF applied to the modified data.

It is evident that Theorem 3 would continue to hold under weaker assumptions. For example, if the requirement on $X(i,j)$ were weakened to require only that $\mathrm{Prob}\left[X(i,j) > \delta > 0\right] > 1 - \varepsilon$ for some $\varepsilon \geq 0$. Then $\mathrm{Prob}\left[\partial^2 f / \partial X^2(g,h) > 0\right]$ could be made arbitrarily close to 1 for $(g,h) \in D_i$, so that (A.25) held with strict inequality, and Theorem 3 followed, also with strict inequality.

However, finding necessary conditions on the $X(i,j)$ to ensure that Theorem 3 holds does not appear easy. In considering such conditions, one may write (A.40) in the form:

$$\frac{1}{Y}\frac{\partial^2 Y}{\partial X^2(g,h)} = 2\sum_{k=h}^{I-g-1}\left[\frac{1}{T(I-k-1,k)} - \frac{1}{T(I-k-1,k+1)}\right] \mathrm{x}$$
$$\sum_{l=h}^{k}\left[\frac{1}{T(I-l-1,l)} - \frac{1}{T(I-l,l)}\right]$$

so that, when the $X(i,j)$ need not be non-negative, a sufficient condition for the left side to be positive is that

$$1/T(g+1,\, h) < 1/T(g,h) \tag{6.1}$$

and

$$1/T(g,\, h+1) < 1/T(g,h) \tag{6.2}$$

for all (g,h).

It is tempting then to contemplate stochastic versions of the inequalities, such as

$$E\left[1/T(g+1,\, h)\right] < E\left[1/T(g,h)\right] \tag{6.3}$$
$$E\left[1/T(g,\, h+1)\right] < E\left[1/T(g,h)\right] \tag{6.4}$$

or such.

However, the choice of such conditions to lead from Theorem 2 to the proof of (A.25), the key to Theorem 3, is not clear.

7. Conclusion

Theorem 3 shows that, under certain distribution free conditions, **the CLF is biased upward**. A simulation test of prediction bias in the chain ladder and other models was carried out by Stanard (1985). One of his experiments dealt with the case in which the total number of claims in an accident year is a Poisson variate and is multinomially distributed over development years. It may be shown that distinct cells in a row of the claim count triangle are then stochastically independent, and so Theorem 3 applies.

Stanard's simulations did in fact find upward bias in the CLF.

8. Acknowledgements

Professor Klaus Schmidt and Dr Klaus Thomas Hess of Technische Universität Dresden gave generously of their time to improve earlier drafts of this paper. Their thorough review identified one major blunder and produced a number of constructive suggestions, leading to other enhancements.

Dr Alois Gisler read even earlier drafts of this paper and provided a number of valuable comments and insights. The paper possibly would not have occurred without his assistance.

The author also acknowledges helpful comment by Dr Thomas Mack.

References

Hertig, J. (1985) A statistical approach to the IBNR-reserves in marine reinsurance. *ASTIN Bulletin*, **15**, 171-183.

Kallenberg, O. (1997) Foundations of modern probability (2nd ed). Springer.

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

Mack, T. (1993) Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, **23**, 213-221.

Mack, T. (1994) Which stochastic model is underlying the chain ladder method? *Insurance: mathematics and economics*, **15**, 133-138.

Renshaw, A.E. (1989) Chain ladder and interactive modelling (claims reserving and GLIM). *Journal of the Institute of Actuaries*, **116**, 559-587.

Stanard, J.N. (1985) A simulation test of prediction errors of loss reserve estimation techniques. *Proceedings of the Casualty Actuarial Society*, **72**, 124-148.

Taylor, G.C. (1986) Loss reserving in non-life insurance. North-Holland, Amsterdam, Netherlands.

Taylor, G. (2000) Loss reserving: an actuarial perspective. Kluwer Academic Publishers, Dordrecht, Netherlands.

Verrall, R.J. (1989) A state space representation of the chain ladder linear model. *Journal of the Institute of Actuaries*, **116**, 589-609.

Verrall, R.J. (1990) Bayes and empirical bayes estimation for the chain ladder model. *Astin Bulletin*, **20**, 217-243.

Verrall, R.J. (1991) On the estimation of reserves from loglinear models. *Insurance: mathematics and economics*, **10**, 75-80.

https://doi.org/10.2143/AST.33.2.503695 Published online by Cambridge University Press

Appendix

Proof of theorems

Lemma 1. Suppose that

$$y = \prod_{i=1}^{n} f_i(x) \tag{A.1}$$

with $x = (x_1, \ldots, x_m)^T$ and $f_i(x) > 0$ for each i. Then

$$\frac{1}{y}\frac{\partial^2 y}{\partial x_k^2} = \sum_{i=1}^{n} \frac{1}{f_i}\frac{\partial^2 f_i}{\partial x_k^2} + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \frac{1}{f_i f_j}\frac{\partial f_i}{\partial x_k}\frac{\partial f_j}{\partial x_k}.$$

Lemma 2 (derivatives of age-to-age factors).

$$\frac{1}{\hat{\nu}(k)}\frac{\partial \hat{\nu}(k)}{\partial X(g,h)} = \frac{\delta(h=k+1)}{T(I-k-1,k+1)} - \frac{\delta(g<I-k)\,\delta(h \le k)\,C(I-k-1,k+1)}{T(I-k-1,k)\,T(I-k-1,k+1)} \tag{A.2}$$

where $\delta(.)$ is defined as follows:

$$\begin{aligned} \delta(C) &= 1 \text{ for condition } C \text{ true;} \\ &= 0 \text{ for condition } C \text{ false;} \end{aligned} \tag{A.3}$$

and it is understood that (A.2) applies to **past observations** ($g + h \le I$).

Further,

$$\frac{1}{\hat{\nu}(k)}\frac{\partial^2 \hat{\nu}(k)}{\partial X^2(g,h)} = \frac{2\delta(g<I-k)\,\delta(h \le k)\,C(I-k-1,k+1)}{T^2(I-k-1,k)\,T(I-k-1,k+1)} \tag{A.4}$$

where this result is again understood to apply only to past observations.

Proof. By (2.10),

$$\partial T(i,j) / \partial X(g,h) = f_i^T e_g e_h^T f_j\,. \tag{A.5}$$

Write the quantity of interest $\hat{\nu}(k)$ in the abbreviated form $\nu = N/D$ where N and D denote numerator and denominator respectively. If N and D depend on variable X, then

$$\begin{aligned} \frac{1}{\nu}\frac{\partial \nu}{\partial X} &= \frac{\partial \log \nu}{\partial X} = \frac{\partial(\log N - \log D)}{\partial X} \\ &= \frac{1}{N}\frac{\partial N}{\partial X} - \frac{1}{D}\frac{\partial D}{\partial X} \\ &= \frac{1}{ND}\left(D\frac{\partial N}{\partial X} - N\frac{\partial D}{\partial X}\right). \end{aligned} \tag{A.6}$$

Now for $\nu = \hat{\nu}(k)$,

$$N = T(I-k-1, k+1) = f_{I-k-1}^T X f_{k+1} \tag{A.7}$$

$$D = T(I-k-1, k) = f_{I-k-1}^T X f_k \tag{A.8}$$

where the matrix notation of Section 2 has been used.

First derivatives of N and D are given by (A.5). Substitution of these into (A.6) yields

$$\begin{aligned}\frac{1}{\hat{\nu}(k)}\frac{\partial \hat{\nu}(k)}{\partial X(g,h)} &= [T(I-k-1,k)\,T(I-k-1,k+1)]^{-1} \times \\ & f_{I-k-1}^T X\left(f_k f_{k+1}^T - f_{k+1} f_k^T\right) e_h e_g^T f_{I-k-1}\end{aligned} \tag{A.9}$$

Note that

$$\begin{aligned} f_k f_{k+1}^T - f_{k+1} f_k^T &= f_k\left(f_k + e_{k+1}\right)^T - \left(f_k + e_{k+1}\right) f_k^T \\ &= f_k e_{k+1}^T - e_{k+1} f_k^T. \end{aligned} \tag{A.10}$$

Substitute (A.10) into (A.9) and recall (2.9) and (2.10) to obtain

$$\begin{aligned}\frac{1}{\hat{\nu}(k)}\frac{\partial \hat{\nu}(k)}{\partial X(g,h)} &= [T(I-k-1,k)\,T(I-k-1,k+1)]^{-1} \times \\ & \left[T(I-k-1,k)\, e_{k+1}^T - C\,(I-k-1,k+1) f_k^T\right] e_h e_g^T f_{I-k-1}\end{aligned} \tag{A.11}$$

Note that

$$f_k^T e_h = \delta\,(h \le k). \tag{A.12}$$

By means of this and similar relations, (A.11) reduces to:

$$\begin{aligned}\frac{1}{\hat{\nu}(k)}\frac{\partial \hat{\nu}(k)}{\partial X(g,h)} &= [T(I-k-1,k)\,T(I-k-1,k+1)]^{-1}\,\delta\,(g \le I-k-1) \\ & \times [\delta\,(h=k+1)\,T(I-k-1,k) - \delta\,(h \le k)\,C\,(I-k-1,k+1)]\end{aligned} \tag{A.13}$$

Note that

$$\begin{aligned}&\delta\,(g \le I-k-1)\,\delta\,(h=k+1) \\ &= \delta\,(g \le I-h)\,\delta\,(h=k+1) = \delta\,(g+h \le I)\,\delta\,(h=k+1) = \delta\,(h=k+1)\end{aligned} \tag{A.14}$$

under the conditions of the lemma.

Substituting (A.14) into (A.13) yields (A.2) as required.

https://doi.org/10.2143/AST.33.2.503695 Published online by Cambridge University Press

To prove (A.4), return to (A.6). In this abbreviated notation there, take a second derivative:

$$\frac{\partial}{\partial X}\left(\frac{1}{\nu}\frac{\partial \nu}{\partial X}\right) = \frac{1}{\nu}\frac{\partial^2 \nu}{\partial X^2} - \left(\frac{1}{\nu}\frac{\partial \nu}{\partial X}\right)^2 \tag{A.15}$$

Now (A.6) yields

$$\frac{\partial}{\partial X}\left(\frac{1}{\nu}\frac{\partial \nu}{\partial X}\right) = \frac{1}{N}\frac{\partial^2 N}{\partial X^2} - \left(\frac{1}{N}\frac{\partial N}{\partial X}\right)^2 - \frac{1}{D}\frac{\partial^2 D}{\partial X^2} + \left(\frac{1}{D}\frac{\partial D}{\partial X}\right)^2 \tag{A.16}$$

By (A.5), (A.7) and (A.8), $\partial^2 N / \partial X^2 = \partial^2 D / \partial X^2 = 0$, so that (A.16) reduces to

$$\begin{aligned}\frac{\partial}{\partial X}\left(\frac{1}{\nu}\frac{\partial \nu}{\partial X}\right) &= \left(\frac{1}{D}\frac{\partial D}{\partial X} - \frac{1}{N}\frac{\partial N}{\partial X}\right)\left(\frac{1}{D}\frac{\partial D}{\partial X} + \frac{1}{N}\frac{\partial N}{\partial X}\right) \\ &= -\left(\frac{1}{\nu}\frac{\partial \nu}{\partial X}\right)\left(\frac{1}{D}\frac{\partial D}{\partial X} + \frac{1}{N}\frac{\partial N}{\partial X}\right)\end{aligned} \tag{A.17}$$

by (A.6).

Substitution of (A.17) in (A.15), and use of (A.6) again yields

$$\frac{1}{\nu}\frac{\partial^2 \nu}{\partial X^2} = -2\left(\frac{1}{\nu}\frac{\partial \nu}{\partial X}\right)\left(\frac{1}{D}\frac{\partial D}{\partial X}\right) \tag{A.18}$$

By (A.8),

$$\partial D / \partial X\,(g,h) = f_{I-k-1}^T e_g e_h^T f_k = \delta(g < I-k)\,\delta(h \le k) \tag{A.19}$$

Substitute (A.2) and (A.19) into (A.18) and note that the term involving $\delta(h = k+1)$ vanishes because $\delta(h = k+1)\,\delta(h \le k) = 0$.

The result is (A.4), as required. □

Lemma 3 (multivariate Jensen inequality). Let $X = (X_1, \ldots, X_m)^T$ where the X_k are stochastically independent random variables. Let $f: R^m \to R$ be twice differentiable in all its arguments, and suppose that

$$\partial^2 f(X) / \partial X_k^2 \ge 0 \text{ for all } X \text{ and for } k = 1, 2, \ldots, m. \tag{A.20}$$

Then

$$E[f(X)] \ge f(E[X]). \tag{A.21}$$

If strict inequality holds in (A.20) for at least one k, and X_k is not degenerate, then strict inequality holds in (A.21).

Proof. An elegant proof of (A.21) appears in Kallenberg (1997, p. 49). It is reproved below in order to obtain the strict inequality.

Let $\mu = (\mu_1, \ldots, \mu_m)^T = E[X]$. Expand $f(X)$ as the Taylor series:

$$f(X) = f(X_1, \ldots, X_{m-1}, \mu_m) + (X_m - \mu_m)\,\partial f(X_1, \ldots, X_{m-1}, \mu_m) / \partial X_m \\ + \tfrac{1}{2}(X_m - \mu_m)^2 \partial^2 f(X_1, \ldots, X_{m-1}, \xi_m) / \partial X^2{}_m \tag{A.22}$$

where $\xi_m = \mu_m + \theta_m (X_m - \mu_m)$ for some $0 < \theta_m < 1$.

Now expand $f(X_1, \ldots, X_{m-1}, \mu_m)$ similarly, then $f(X_1, \ldots, X_{m-2}, \mu_{m-1}, \mu_m)$, and so on to obtain

$$f(X) = f(\mu) + \sum_{k=1}^{m} (X_k - \mu_k)\,\partial f(X_1, \ldots, X_{k-1}, \mu_k, \ldots, \mu_m) / \partial X_k \\ + \sum_{k=1}^{m} \tfrac{1}{2}(X_k - \mu_k)^2 \partial^2 f(X_1, \ldots, X_{k-1}, \xi_k, \mu_{k+1}, \ldots, \mu_m) / \partial X_k^2. \tag{A.23}$$

Take expectations on both sides of (A.23). Note that

$$E\left[(X_k - \mu_k)\,\partial f(X_1, \ldots, X_{k-1}, \mu_k, \ldots, \mu_m) / \partial X_k\right] \\ = E(X_k - \mu_k)\, E\left[\partial f(X_1, \ldots, X_{k-1}, \mu_k, \ldots, \mu_m) / \partial X_k\right] \\ = 0 \tag{A.24}$$

where the middle step follows from the stochastic independence of X_k from $(X_1, \ldots, X_{k-1})$.

By (A.20),

$$E\left[(X_k - \mu_k)^2 \partial^2 f(X_1, \ldots, X_{k-1}, \xi_k, \mu_{k+1}, \ldots, \mu_m) / \partial X_k^2\right] \geq 0 \tag{A.25}$$

with strict inequality if strict inequality holds in (A.20) and X_k is not degenerate.

The lemma then follows. □

Proof of Theorem 2. Consider $Y(i)$ defined by (5.1), with $\hat{v}(k)$ defined by (3.1) and (2.3). The observations $X(g,h)$ involved in the $\hat{v}(k)$ constituting $Y(i)$ are just those in D_i. This justifies (5.2).

Now consider $(g,h) \in D_i$. Note that Lemma 1 is applicable to $Y(i)$ because of (5.1). Hence

https://doi.org/10.2143/AST.33.2.503695 Published online by Cambridge University Press

$$\frac{1}{Y(i)}\frac{\partial^2 Y(i)}{\partial X^2(g,h)} = \sum_{k=I-i}^{I-1}\frac{1}{\hat{\nu}(k)}\frac{\partial^2\hat{\nu}(k)}{\partial X^2(g,h)} + \sum_{\substack{k,l=I-i\\k\neq l}}^{I-1}\frac{1}{\hat{\nu}(k)\hat{\nu}(l)}\frac{\partial\hat{\nu}(k)}{\partial X(g,h)}\frac{\partial\hat{\nu}(l)}{\partial X(g,h)}. \tag{A.26}$$

In this equation $I-i \le k \le I-1$. Since $(g,h) \in D_i$ it is also the case that $h-1 \le k \le I-1-g$. This yields $\max\{I-i, h-1\} \le k \le I-1-g$.

Therefore, there are two cases to be considered:

- In the case $h \le I-i$ one has $I-i \le k \le I-1-g$
- In the case $h \ge I-i+1$ one has $h-1 \le k \le I-1-g$.

Case I: $h \le I-i$.

In this case

$$g < I-k, \quad h \le k. \tag{A.27}$$

Under these conditions, combination of (3.1) with (A.2) gives

$$\partial\hat{\mathrm{v}}(k)/\partial X(g,h) = -C(I-k-1, k+1)/T^2(I-k-1, k). \tag{A.28}$$

Similarly, combination of (3.1) with (A.4) gives

$$\partial^2\hat{\mathrm{v}}(k)/\partial X^2(g,h) = 2C(I-k-1, k+1)/T^3(I-k-1, k). \tag{A.29}$$

Substitution of (3.1), (A.28) and (A.29) into (A.26) yields

$$\begin{aligned}
\frac{1}{Y(i)}\frac{\partial^2 Y(i)}{\partial X^2(g,h)} &= 2\sum_{k=I-i}^{I-g-1}\frac{C(I-k-1,k+1)}{T(I-k-1,k+1)\,T^2(I-k-1,k)} \\
&+ 2\sum_{k=I-i}^{I-g-1}\sum_{l=I-i}^{k-1}\frac{C(I-k-1,k+1)\,C(I-l-1,l+1)}{T(I-k-1,k+1)\,T(I-k-1,k)\,T(I-l-1,l+1)\,T(I-l-1,l)} \\
&= 2\sum_{k=I-i}^{I-g-1}\frac{C(I-k-1,k+1)}{T(I-k-1,k+1)\,T(I-k-1,k)} \\
&\times\left[\frac{1}{T(I-k-1,k)} + \sum_{l=I-i}^{k-1}\frac{C(I-l-1,l+1)}{T(I-l-1,l+1)\,T(I-l-1,l)}\right]
\end{aligned} \tag{A.30}$$

Note that Assumption 3 implies that $T(g,h) > 0$ for $(g,h) \in D_i$, and so guarantees the existence of all ratios in (A.30).

Note also that, by (2.2), the second member within the square bracket may be expanded as follows:

$$\frac{C(I-l-1,l+1)}{T(I-l-1,l+1)\,T(I-l-1,l)} = \frac{1}{T(I-l-1,l)} - \frac{1}{T(I-l-1,l+1)} \tag{A.31}$$

Substitute (A.31) into the square bracket in (A.30) to obtain

$$\begin{aligned}
&\frac{1}{T(I-k-1,k)} + \sum_{l=I-i}^{k-1} \frac{1}{T(I-l-1,l)} - \sum_{l=I-i+i}^{k} \frac{1}{T(I-l,l)} \\
&= \frac{1}{T(i-1,I-i)} + \sum_{l=I-i+1}^{k} \left[\frac{1}{T(I-l-1,l)} - \frac{1}{T(I-l,l)} \right] \\
&= \frac{1}{T(i-1,I-i)} + \sum_{l=I-i+1}^{k} \frac{R(I-l,l)}{T(I-l-1,l)\,T(I-l,l)}
\end{aligned} \tag{A.32}$$

by (2.1).

Substitute (A.32) for the square bracket in (A.30) to obtain

$$\begin{aligned}
\frac{1}{Y(i)} \frac{\partial^2 Y(i)}{\partial X^2(g,h)} = 2 \sum_{k=I-i}^{I-g-1} \frac{C(I-k-1,k+1)}{T(I-k-1,k+1)(I-k-1,k)} \times \\
\left[\frac{1}{T(i-1,I-i)} + \sum_{l=I-i+1}^{k} \frac{R(I-l,l)}{T(I-l-1,l)\,T(I-l,l)} \right].
\end{aligned} \tag{A.33}$$

This proves (5.3).

Case II: $h > I-i$.

It follows from the argument immediately preceding the proof of Case I that

$$g < I-k, \quad k \geq h-1 \tag{A.34}$$

(A.34) may be written as the two sub-cases:

$k = h-1$

and

$g < I-k,\ h \leq k.$

Then Lemma 2 gives

$$\frac{\partial \hat{\nu}(h-1)}{\partial X(g,h)} = \frac{1}{T(I-h,h-1)} \tag{A.35}$$

$$\frac{\partial^2 \hat{\nu}(h-1)}{\partial X^2(g,h)} = 0 \tag{A.36}$$

$$\frac{\partial \hat{\nu}(k)}{\partial X(g,h)} = -\frac{C(I-k-1,k+1)}{T^2(I-k-1,k)}, \quad k \geq h,\ g \leq I-k-1 \tag{A.37}$$

$$\frac{\partial^2 \hat{\nu}(k)}{\partial X^2(g,h)} = \frac{2C(I-k-1,k+1)}{T^3(I-k-1,k)},\ k \geq h,\ g \leq I-k-1. \tag{A.38}$$

Substitute (A.35) – (A.38) into (A.26) to obtain

$$\begin{aligned}
\frac{1}{Y(i)} \frac{\partial^2 Y(i)}{\partial^2 X(g,h)} &= \frac{1}{\hat{\nu}(h-1)} \frac{\partial^2 \hat{\nu}(h-1)}{\partial X^2(g,h)} + \sum_{k=h}^{I-g-1} \frac{1}{\hat{\nu}(k)} \frac{\partial^2 \hat{\nu}(k)}{\partial X^2(g,h)} \\
&+ \sum_{\substack{k,l=h \\ k \neq l}}^{I-g-1} \frac{1}{\hat{\nu}(k)\hat{\nu}(l)} \frac{\partial \hat{\nu}(k)}{\partial X(g,h)} \frac{\partial \hat{\nu}(l)}{\partial X(g,h)} \\
&+ \frac{2}{\hat{\nu}(h-1)} \frac{\partial \hat{\nu}(h-1)}{\partial X(g,h)} \sum_{k=h}^{I-g-1} \frac{1}{\hat{\nu}(k)} \frac{\partial \hat{\nu}(k)}{\partial X(g,h)} \\
&= 2 \sum_{k=h}^{I-g-1} \frac{C(I-k-1,k+1)}{T(I-k-1,k+1)\,T^2(I-k-1,k)} \\
&+ 2 \sum_{k=h}^{I-g-1} \frac{C(I-k-1,k+1)}{T(I-k-1,k+1)\,T(I-k-1,k)} \sum_{l=h}^{k-1} \frac{C(I-l-1,l+1)}{T(I-l-1,l+1)\,T(I-l-1,l)} \\
&- \frac{2}{T(I-h,h)} \sum_{k=h}^{I-g-1} \frac{C(I-k-1,k+1)}{T(I-k-1,k+1)\,T(I-k-1,k)}.
\end{aligned} \tag{A.39}$$

As in Case I, Assumption 3 guarantees the existence of all the ratios in (A.39).

Now apply (A.31) and use the same mode of calculation as led from (A.30) to (A.33). Then (A.39) becomes:

$$\begin{aligned}
\frac{1}{Y(i)} \frac{\partial^2 Y(i)}{\partial X^2(g,h)} &= 2 \sum_{k=h}^{I-g-1} \frac{C(I-k-1,k+1)}{T(I-k-1,k+1)\,T(I-k-1,k)} \times \\
&\quad \sum_{l=h}^{k} \frac{R(I-l,l)}{T(I-l-1,l)\,T(I-l,l)}.
\end{aligned} \tag{A.40}$$

This proves (5.4). □

Proof of Theorem 3. Consider $Y(i)$ defined by (5.1). By Theorem 2,

$\partial^2 Y(i) / \partial X^2(g,h) \geq 0$ for all (g,h)

with strict inequality for some (g,h), namely those in D_i. It follows from a multivariate form of Jensen's inequality (see Lemma 3) that

$$E[Y(i)] > \bar{Y}(i) = \prod_{k=I-i}^{I-1} \bar{\nu}(k) \tag{A.41}$$

https://doi.org/10.2143/AST.33.2.503695 Published online by Cambridge University Press

where $\bar{Y}(i)$ is the value obtained by replacing each $X(g,h)$ in $Y(i)$ by its expectation, and $\bar{v}(k)$ is similarly defined.

By (3.1),

$$\begin{aligned}\bar{v}(k) &= E[T(I-k-1,k+1)] \,/\, E[T(I-k-1,k)] \\ &= \sum_{g=0}^{I-k-1} E[R(g,k+1)] \,/\, \sum_{g=0}^{I-k-1} E[R(g,k)] \quad \text{[by (2.3)]} \\ &= \eta(k), \quad \text{by (4.5)}.\end{aligned} \tag{A.42}$$

Substitute (A.42) in (A.41):

$$E[Y(i)] > \prod_{k=I-i}^{I-1} \eta(k) = E[R(i,I)] \,/\, E[R(i,I-i)] \tag{A.43}$$

by (4.5). This proves (5.5).

Now take expectations on both sides of (3.2):

$$E\left[\hat{R}(i,I)\right] = E[R(i,I-i)]\,E[Y] > E[R(i,I)], \quad \text{by (A.43)}. \tag{A.44}$$

The first step leading to (A.44) is justified as follows. $Y(i)$ is defined by (5.1) and $\hat{v}(k)$ by (3.1), which shows that the rows of X involved in $Y(i)$ are $0, 1, \ldots, i-1$. Thus, $R(i,I-i)$ and $Y(i)$ are stochastically independent.

□

Greg Taylor
Taylor Fry Consulting Actuaries
Level 8, 30 Clarence Street
Sydney NSW 2000
Australia

Professorial Associate,
Centre for Actuarial Studies
Faculty of Economics and Commerce
University of Melbourne
Parkville VIC 3052
Australia

Phone: 61-2-9249-2901
Fax: 61-2-9249-2999
greg@taylorfry.com.au

https://doi.org/10.2143/AST.33.2.503695 Published online by Cambridge University Press