The Variability of the IBNR – Mack, Murphy and Peterson Formulas.

By Thomas G. Kabele

Increasingly actuaries are called on to estimate the variability of the loss reserves In this note we examine estimates of the mean square error of the chain ladder estimate of Loss reserves including

(1) Thomas Mack's estimates (1993, 1994, 1999)
(2) Timothy Peterson (1980) estimate.
(3) Daniel Murphy's estimates (1994)

Section 1 Formulas.
We give the classical formulas for the loss reserves.

Section 2 Consistency.
We prove that the Mack assumptions are consistent for all his formulas including the $\alpha = 0$ (simple average), $\alpha = 1$ (weighted average) and $\alpha = 2$ (weighted average) formulas. We show that the Murphy assumptions are consistent for the simple average case but not for the other cases. We show that the link factors are not independent for the $\alpha = 1$ and $\alpha = 2$ cases. (Greg Taylor mentioned the independent assumption in his proof of Mack's $\alpha = 1$ case but the proof doesn't need independence.) The link factors might be independent in the α =0 case.

Section 3 Mack and L-predictors.
Mack completed the proof for $\alpha = 1$ case and outlined the proof for the other cases. In the second part of the paper we completed those and produced some new estimates, which we call "L-predictors." The "L-predictors" give practically the same result in the typically case when the claims data increases by age, but a different result when the claims data decreases (as with case reserves). We also proved a "summation theorem" which was needed to plug a potential problem with the proofs.

Section 4 - Peterson.
Timothy Peterson (1980) noted the variability of loss reserves might be estimated using maximum and minimum link ratios, instead of just averages -- but he noted that this method was flawed. We showed how to refine his technique by using the standard deviations of the link factors.

Section 5 -Murphy.
We redid Murphy's proof for the simple average case using a notation consistent with Mack's. Murphy did not use the mean square error to estimate the variability, but defined the variability in terms of the "process risk" and the "parameter risk."

Section 1. NOTATION for the Chain Ladder Formulas.
1. We assume our claims data is grouped by accident year and development age. (We could also use report years or policy years in place of accident years and use valuation date instead of development age.) We let $m = I =$ number of accident years; $n = J =$ number of development age intervals. We assume $m \geq n$.

2. We let $C_{i,k}$ = cumulative claims for accident year $i, (1 \le i \le m)$ up to development age $k, (1 \le k \le n)$. We assume $i = 1$ is the oldest accident year, and development age $k = 1$ is the first development period. The claims could be in monetary units (say $) and could be paid, case reserves, or case incurred (case reserve + paid); or they could represent claim counts (reported, outstanding, or closed with payment).
3. We assume that D = $\{C_{i,k} : i + k \le m+1; 1 \le k \le n\}$ = "known" data, and that we are to estimate the "unknown" part of the rectangle, namely $\{C_{i,k} : i + k > m+1\}$. We especially want the "ultimate" values at development age n, namely $ULT = \{C_{i,n} : i = 1,..., m\}$.

In the following claims rectangle there are m=5 accident years, n=4 development years and the "known" part of the triangle is listed.

$$\begin{pmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,n} \\ C_{2,1} & C_{2,2} & C_{2,3} & C_{2,n} \\ C_{3,1} & C_{3,2} & C_{3,3} & \\ C_{4,1} & C_{4,2} & & \\ C_{m,1} & & & \end{pmatrix}$$

4. An accident year is said to be "fully developed" at development age k if $C_{i,k} = C_{i,k+1} = C_{i,k+2}$ etc. We assume that the oldest accident years with $i \le 1 + m - n$ are fully developed, and that $q = 2 + m - n$ is the first accident year that is not fully developed.

S1.1. CHAIN LADDER ESTIMATES

The following definitions are from Mack, Cas. Forum, 1994, p113 and Astin Bulletin 1999. Using only the known part of the loss rectangle we compute factors (1)-(6) below. We are given some weights: $\{w_{i,j} : i + j \le m;\ 1 \le j \le n-1\}$ (often equal to 1). We then define:

(1) Age to age factors: $F_{i,k} = C_{i,k+1} / C_{i,k}\ (i + k \le m)\ (k = 1,..., n-1)$

(2) Average age to age factors

(a) $\hat{f}_k = \sum_{i=1}^{m-k} F_{i,k} w_{i,k} / \sum_{i=1}^{m-k} w_{i,k}\ \ (k = 1,..., n-1)$ (simple average, $a = 0$)

(b) $\hat{f}_k = \sum_{i=1}^{m-k} F_{i,k} C_{i,k} w_{i,k} / \sum_{i=1}^{m-k} C_{i,k} w_{i,k}\ \ (k = 1,..., n-1)$ (weighted average, $a = 1$)

(c) $\hat{f}_k = \sum_{i=1}^{m-k} F_{i,k} C_{i,k}^2 w_{i,k} / \sum_{i=1}^{m-k} C_{i,k}^2 w_{i,k}\ \ (k = 1,..., n-1)$ (least squares average, $a = 2$)

(d) General case: $\hat{f}_k = \sum_{i=1}^{m-k} F_{i,k} \hat{b}_{i,k} / \hat{b}_k$ where $\hat{b}_{i,k} = w_{i,k} C_{i,k}^{a}$ for $i + k \le m;\ 1 \le k \le n-1$ and $a \in \{0,1,2\}$ and $\hat{b}_k = \sum_{i=1}^{m-k} \hat{b}_{i,k}$

(3) Age to Ultimate factors: $\hat{u}_i = \hat{f}_{m+1-i/n} = \begin{cases} 1 \text{ if } i \le 1+m-n \\ \hat{f}_{m+1-i} \cdots \hat{f}_{n-1} \text{ if } 1+m-n < i \le m \end{cases}$

(4) Spread Factor $\hat{s}_k^2 = \frac{1}{m-k-1}\sum_{i=1}^{m-k} w_{i,k} C_{i,k}^{a} (F_{i,k} - \hat{f}_k)^2 \quad 1 \le k \le \min(m-2, n-1)$

(5a) The "losses to date" are those that lie on the diagonal $LTD = \{C_{i,j} : i + j = m+1\}$.

(5b) The ultimate losses are those on the last column $ULT = \{C_{i,n} : i = 1,...,m\}$

(5c) The unknown remaining losses are $R_i = ULT_i - LTD_i = C_{i,n} - C_{i,m+1-i}$

(6a) Alternative notation for claims in the known region: $\hat{C}_{i,k} = C_{i,k}$

(6b) Estimated Claims in the unknown region: $\hat{C}_{i,k} = C_{i,p}\, \hat{f}_p \cdots \hat{f}_{k-1}$ where $k > p = m+1-i$

(6c) Estimated Ultimate Claims, accident year i: $\hat{U}LT = \{\hat{C}_{i,n} : i = 1,...,m\}$ where

$\hat{C}_{i,n} = C_{i,p}\, \hat{f}_p \cdots \hat{f}_{n-1}$ where $p = m+1-i$ for $i = 1,....,\min(m,n)$ and

$\hat{C}_{i,n} = C_{i,n}$ for $i > \min(m,n)$.

(6d) Estimated Loss Reserve, accident year i (IBNR): $\hat{R}_i = \hat{U}LT_i - LTD_i$.

(6e) Estimated Loss Reserve, all accident years $\hat{R} = \sum_{i=1}^{m} \hat{R}_i$.

S2. CHAIN LADDER ASSUMPTIONS and Their Consistency.

Mack in the 1993, 1994 and 1999 papers showed how to estimate the "mean square" error of the estimated loss reserves by making certain reasonable stochastic assumptions. These assumptions are shown below, using contingent expectation. The parameter $a \in \{0,1,2\}$ and weights $\{w_{i,j} : 1 \le i \le m; 1 \le j \le n\}$ are fixed. by the actuary.

We assume: that the $\{C_{i,k} : 1 \le i \le m,\ 1 \le k \le n\}$ are random variables on some probability space $(\Omega, \mathcal{A}, \text{Prob})$ and that $C_{i,k} > 0$. We do not require that for each fixed accident year i that $\{C_{i,k} : k = 1,...,n\}$ be an increasing sequence. We assume there are unknown constants $\{f_k : k = 1,...,n-1\}$ and $\{s_k^2 : k = 1,...,n-1\}$ such that:

CL1. $E(F_{i,k} \mid C_{i,1},..., C_{i,k}) = f_k$ where $(i = 1,...,m)(k = 1,...,n-1)$

CL2. $Var(F_{i,k} \mid C_{i,1},..., C_{i,k}) = s_k^2 / b_{i,k}$ where $(i = 1,...,m)(k = 1,...,n-1)$ and

$b_{i,k} = w_{i,k}\, C_{i,k}^{a}$ and where $a \in \{0,1,2\}$.

CL3. The accident years $\{C_{g,1},....,C_{g,n}\}$ and $\{C_{i,1},....,C_{i,n}\}$ are independent for all g and i.

Note that the assumptions and the weights $w_{i,k}$ and $b_{i,k}$ apply to both the known part of the loss rectangle $D = \{C_{i,j} : i + j \le m+1\}$ and the unknown part: $\{C_{i,j} : i + j > m+1\}$.

S2.1. Alternate notation for contingent expectations.
We use an "operator" notation for contingent expectations. Let H be a collection of random variables -- or more generally a s subalgebra of A . We often write $E_H(X)$ in place of $E(X|H)$ where X is any integrable random variable. It is important to note that $E_H(X) = E(X|H)$ is a random variable and not a constant. We define:

(a) $\text{Var}_H(X) = E_H(X^2) - E_H(X)E_H(X)$

(b) " X is perpendicular to Y " $(X \perp Y)$ means $E(XY) = 0$

(c) " X is H-measurable" is as defined in advanced texts, such as Loeve, Probability Theory. If H is a set of random variables, then a sufficient condition for "X is H-measurable" is that X is a linear combination or a continuous function of elements of H.

We list some of the properties of conditional expectation below, where H, G, D, H_i, D_i are various sets of random variables (or s subalgebras of A).

0. Linear combinations.

$$E_H(aX + bY) = a\,E_H(X) + b\,E_H(Y) \text{ where } a,b \in \text{R}$$

1. Iteration.

(a) $EE_H(X) = E(X)$

(b) $E_H E_G(X) = E_G(X) = E_G E_H(X)$ if $G \subseteq H$.

2. Independence.

(a) $E_{H \cup G}(X) = E_G(X)$ if H is independent of both G and X

(b) If G and X are independent, then $E_G(X) = E(X)$ and $E_G(X^2) = E(X^2)$ and $Var_G(X) = Var(X)$

(c) Suppose that $X_i : S \to$ □ are H_i measurable, and H_i are independent and suppose that $D_i \subseteq H_i$. Then $E_{D_i}(X_i)$ are independent and $E_{D_1 \cup D_2}(X_1 X_2) = E_{D_1}(X_1)\,E_{D_2}(X_2) = E_{D_1 \cup D_2}(X_1)\,E_{D_1 \cup D_2}(X_2)$

3. Variance

(a) $\text{Var}(X) = E\,\text{Var}_H(X) + \text{Var}\,E_H(X)$

(b) $\text{Var}_G(X) = E_G\,\text{Var}_H(X) + \text{Var}_G\,E_H(X)$ if $G \subseteq H$.

4. Factoring out.
If h is an H-measurable random variable, and X is any (integrable) random variable, then

(a) $E_H(X\,h) = h\,E_H(X)$ and $Var_H(X\,h) = h^2 Var_H(Y)$.

(b) $E_H(X + h) = h + E_H(X)$ and $Var_H(X + h) = Var_H(X)$

5. Other Properties.

(a) $E_H(X) = X$ if X is H-measurable

(b) $E_H(X) = 0$ if X is perpendicular to the H-measurable functions

We can prove all of the above using the following properties:

(a) $E_H(X)$ is "H-measurable"

(b) $X - E_H(X)$ is "perpendicular" to the "H-measurable" random variables.

See Loeve and other advanced texts on probability for details.

S2.2 SETS of VARIABLES and the CHAIN LADDER ASSUMPTIONS.

We use the following sets of variables in the formulas in this paper.

$D = \{C_{i,k} : i + k \le m + 1\}$ = known values

$B_k = \{C_{i,j} : j \le k\}$ = data on or prior to development age k.

$A_i = \{C_{i,k} : k = 1,...,n\}$ = i-th accident year.

$D_i = D \cap A_i$ = i-th accident year, but known values.

$G_{i,k} = A_i \cap B_k = \{C_{i,j} : j \le k\}$ = i-th accident year, on or prior to development age k.

${}_1A_k = A_1 \cup A_2 \cup \cdots \cup A_k =$ 1st to kth accident year

$L_k = B_k \cap {}_1A_{m-k} = \{C_{i,j} : 1 \le i \le m-k; 1 \le j \le k\} \subset D \cap B_k$

The $\{B_k, A_i, G_{i,k}\}$ include both known and unknown values.

Using the above notation for conditional expectation the Chain Ladder hypotheses can be defined as follows:

Chain Ladder Assumptions (see Mack 1999, page 362)

CL1. $E_{G_{i,k}}(F_{i,k}) = f_k$ where $(i = 1,...,m)(k = 1,...,n-1)$

CL2. $Var_{G_{i,k}}(F_{i,k}) = \boldsymbol{s}_k^2 / \boldsymbol{b}_{i,k}$ where $\boldsymbol{b}_{i,k} = w_{i,k}\, C_{i,k}^{a}$ $(i = 1,...,m)(k = 1,...,n-1)$ and

$\boldsymbol{b}_{i,k} = w_{i,k}\, C_{i,k}^{a}$ and where $\boldsymbol{a} \in \{0,1,2\}$.

CL3. The accident years A_i are independent for all i.

Other Formulations

Let us define the incremental claims by:

$S_{i,k+1} = C_{i,k+1} - C_{i,k}$ for $S_{i,k+1} = C_{i,k+1} - C_{i,k}$ and $S_{i,1} = C_{i,1}$.

We can give the following alternative versions of CL1 and CL2.

For cumulative claims:

CL1: $E_{G_{i,k}}(C_{i,k+1}) = C_{i,k}\ f_k$.

CL2: $Var_{G_{i,k}}(C_{i,k+1}) = \boldsymbol{s}_k^2 C_{i,k}^2 / \boldsymbol{b}_{i,k}$ where $\boldsymbol{b}_{i,k} = w_{i,k}\, C_{i,k}^{a}$.

For Incremental Claims:

CL1: $E_{G_{i,k}}(S_{i,k+1}) = C_{i,k}\,(f_k - 1)$.

CL2: $Var_{G_{i,k}}(S_{i,k+1}) = \boldsymbol{s}_k^2 C_{i,k}^2 / \boldsymbol{b}_{i,k}$ where $\boldsymbol{b}_{i,k} = w_{i,k}\, C_{i,k}^{a}$.

Written for and presented at 8th GCA, Mumbai 10-11 March, 2006

The formulas are summarized below:

	Link factors	Cumulative Claims	Incremental Claims
CL1	$E_{G_{i,k}}(F_{i,k}) = f_k$	$E_{G_{i,k}}(C_{i,k+1}) = C_{i,k}\ f_k$	$E_{G_{i,k}}(S_{i,k+1}) = C_{i,k}\ (f_k - 1)$
CL2 $a = 0$	$Var_{G_{i,k}}(F_{i,k}) = s_k^2 / w_{i,k}$	$Var_{G_{i,k}}(C_{i,k+1}) = s_k^2 C_{i,k}^2 / w_{i,k}$	$Var_{G_{i,k}}(S_{i,k+1}) = s_k^2 C_{i,k}^2 / w_{i,k}$
CL2 $a = 1$	$Var_{G_{i,k}}(F_{i,k}) = s_k^2 / w_{i,k} C_{i,k}$	$Var_{G_{i,k}}(C_{i,k+1}) = s_k^2 C_{i,k} / w_{i,k}$	$Var_{G_{i,k}}(S_{i,k+1}) = s_k^2 C_{i,k} / w_{i,k}$
CL2 $a = 2$	$Var_{G_{i,k}}(F_{i,k}) = s_k^2 / w_{i,k} C_{i,k}^2$	$Var_{G_{i,k}}(C_{i,k+1}) = s_k^2 / w_{i,k}$	$Var_{G_{i,k}}(S_{i,k+1}) = s_k^2 / w_{i,k}$

Remark. If the Claims $C_{i,j}$ are in monetary units (say $) and if $a = 1$, then the s_k^2 and the $b_{i,k} = w_{i,k}\, C_{i,k}^a$ will be expressed in dollars. If $a = 2$, then s_k^2 and the $b_{i,k}$ will be expressed in dollars squared. If $a = 0$, then s_k^2 and the $b_{i,k}$ are dimensionless. In all cases the quotient $s_k^2 / b_{i,k}$ is dimensionless.

Remark. If $a = 0$ then $Var_{G_{i,k}}(F_{i,k}) = s_k^2 / w_{i,k}$ is a constant which does not depend on the claim amounts. Also

$$E_{G_{i,k}}(F_{i,k}^2) = \mathrm{Var}_{G_{i,k}}(F_{i,k}^2) + E_{G_{i,k}}(F_{i,k})^2 = s_k^2 / w_{i,k} + f_k^2$$

S2.3. CONSISTENCY

Before one uses any model one should prove that the model is consistent. We will therefore give examples of random variables that satisfy each of the three assumptions -- for $a \in \{0,1,2\}$. We use the following theorem:

Theorem (from the Kolmogorov Existence Theorem). Given any distribution functions $\{F_n(x) : 1 \le n < \infty\}$ we can find a probability space $(\Omega, \mathcal{A}, \mathrm{Prob})$ and independent random variables $\{X_n : 1 \le n < \infty\}$ having the $\{F_n(x) : 1 \le n < \infty\}$ as distributions.

Proof. See Billingsley, Probability and Measure (Wiley, 3rd ed, 1995) page 265 and cf. 486, 73.

Remark. To describe the model we need only describe only one of the rows, for suppose the first row satisfies CL1 and CL2. By the Kolmogorov Existence Theorem we can find for each row $i = 2,...,m$ variables $\{C_{i,1}, \ldots, C_{i,n}\}$ which are independent of the other rows and have the same distribution as the first row.

Theorem. The models are all consistent-- that is we can define positive random variables that satisfy CL1-CL2-CL3 for $a \in \{0,1,2\}$. .

Proof for $a = 0$ model. Fix the i-th accident year. By the Kolmogorov existence theorem pick independent random variables $\{T_0,...,T_{n-1}\}$ for which $E(T_k) = f_k$ and $\mathrm{Var}(\mathrm{T_k}) = s_k^2 / w_{i,k}$. Let $C_{i,1} = T_0$ and $C_{i,k+1} = C_{i,k}\,T_k$. Then

$F_{i,k} = T_k$ for $k=1,..., n-1$. By computation we find

$$E_{G_{i,k}}(F_{i,k}) = E(T_k \mid C_{i,1},..., C_{i,k}) = E(T_k \mid T_0,...,T_{k-1})$$

But since the $\{T_k\}$ are independent the above equals $E(T_k) = f_k$.

Likewise

$$Var_{G_{i,k}}(F_{i,k}) = Var(T_k \mid C_{i,1},..., C_{i,k}) = Var(T_k \mid T_0,...,T_{k-1})$$

$$= Var(T_k) = s_k^2 / w_{i,k}$$

Proof for $a = 2$ model. Let $\{\boldsymbol{e}_k : 1 \le k \le n\}$ be independent random variables with mean $E(\boldsymbol{e}_k) = 0$ and variance $Var(\boldsymbol{e}_k) = s_k^2$. Let $C_{i,1}$ be any positive random variable independent of the $\{\boldsymbol{e}_k : 1 \le k \le n\}$. Define

$$C_{i,k+1} = f_k\, C_{i,k} + \boldsymbol{e}_{k+1}$$

By computations, since $C_{i,k}$ is $G_{i,k}$ measurable:

$$E_{G_{i,k}}(C_{i,k+1}) = E(C_{i,k+1} \mid C_{i,1},\ldots, C_{i,k}) = f_k C_{i,k} + E_{G_{i,k}}(\boldsymbol{e}_{i,k+1})$$

But by independence $E_{G_{i,k}}(\boldsymbol{e}_{i,k+1}) = E(\boldsymbol{e}_{i,k+1}) = 0$. Likewise

$$Var_{G_{i,k}}(C_{i,k+1}) = Var(C_{i,k+1} \mid C_{i,1},\ldots,C_{i,k}) = Var_{G_{i,k}}(\boldsymbol{e}_{i,k+1})$$

Again by independence $Var_{G_{i,k}}(\boldsymbol{e}_{i,k+1}) = Var(\boldsymbol{e}_{i,k+1}) = s_k^2$

Proof for $a = 1$. Let $\{\boldsymbol{e}_k : 1 \le k \le n\}$ and $C_{i,1}$ be as in the $a = 2$ model. Define

$$C_{i,k+1} = f_k\, C_{i,k} + \boldsymbol{e}_{k+1}\sqrt{C_{i,k}}$$

Then

$$E_{G_{i,k}}(C_{i,k+1}) = E(C_{i,k+1} \mid C_{i,1},\ldots, C_{i,k}) = f_k C_{i,k} + \sqrt{C_{i,k}}\,E_{G_{i,k}}(\boldsymbol{e}_{i,k+1})$$

$$= f_k C_{i,k}$$

Likewise

$$Var_{G_{i,k}}(C_{i,k+1}) = Var(C_{i,k+1} \mid C_{i,1},\ldots,C_{i\,k}) = C_{i,k}\,Var_{G_{i,k}}(\boldsymbol{e}_{i,k+1})$$

$$= C_{i,k} Var(\boldsymbol{e}_{i,k+1}) = C_{i,k} s_k^2$$

The proof is done.

Remark. The above proof shows that for $a = 0$ we can add an additional hypothesis:

CL4. The $\{F_{i,k} : 1 \le k \le n\}$ are independent.

We cannot, however, extend CL4 to the $a = 1$ or $a = 2$ cases. (By some additional work we can find cases where CL1-CL2-CL3 are true but CL4 is false for $a = 0$.)

Proposition. Assume that all of our random variables are based on a probability space $(\Omega, \mathcal{A}, \text{Prob})$. Fix the $i-th$ accident year. Assume none of the random variables $\{\{C_{i,k} : 1 \le k \le n\}$ are constant. Let $F_{0,k} = C_{i,1}$. We have:

(a) In the $\alpha = 1$ or $\alpha = 2$ cases the link ratios $\{F_{i,k} : 0 \le k \le n\}$ cannot be independent.

(b) In all cases the cumulative claims $\{C_{i,k} : 1 \le k \le n\}$ cannot be independent.

(c) In all cases the incremental claims $\{S_{i,k} : 1 \le k \le n\}$ cannot be independent.

Proof (a) Note that $\{C_{i1},...,C_{i,k}\}$ and $\{F_{i,0},...,F_{i,k-1}\}$ generate the same σ-subalgebra of $\mathcal{A}$ since $F_{i,k} = C_{i,k+1} / C_{i,k}$. If the $\{F_{i,k} : 1 \le k \le n\}$ were independent then $Var(F_{i,k} \mid C_{i,1}, \ldots C_{i,k}) = Var(F_{i,k} \mid F_{i,0}, \ldots F_{i,k-1})$ would be constant, but this is not true by CL2 for the $\alpha = 1$ or $\alpha = 2$ cases.

Proof (b) If (for fixed i) the $\{C_{i,k} : 1 \le k \le n\}$ were independent then $E(C_{i,k+1} \mid C_{i,1}, \ldots, C_{i,k})$ would be constant, but this is not true by CL1.

Proof (c) Note that $\{C_{i,k} : 1 \le k \le n\}$ and $\{S_{i,k} : 1 \le k \le n\}$ generate the same σ-subalgebra of $\mathcal{A}$ since $C_{i,k+1} = C_{i,k} + S_{i,k+1}$ and $C_{i,1} = S_{i,1}$. If (for fixed i) the $\{S_{i,k} : 1 \le k \le n\}$ were independent then $E(S_{i,k+1} \mid S_{i,1}, \ldots, S_{i,k})$ would be constant, but this is not true by CL1.

Remark. If we replace CL1 by the CL1A below, then we can add CL5 below - i.e. we can make the $m \times n$ incremental claims $\{S_{i,j} : 1 \le i \le m, 1 \le j \le n\}$ mutually independent.

CL1A. For all i: $E(S_{i,k+1} \mid S_{i,1}, \ldots, S_{i,k}) = s_{k+1}$ where s_k is constant.

CL5. The $\{S_{i,j} : 1 \le i \le m, 1 \le j \le n\}$ are independent .

Proof. By the Kolmogorov existence theorem choose $\{S_{i,j} : 1 \le i \le m, 1 \le j \le n\}$ to be independent with $Var(S_{i,k+1}) = \sigma_k^2$ and $E(S_{i,k}) = s_k$ (for all i and k). Then CL5 holds. Also CL1A holds by independence of the $\{S_{i,j} : 1 \le i \le m, 1 \le j \le n\}$.

Definition. We define the estimator $\hat{\sigma}_k^2$ of the parameter σ_k^2 by

$$\hat{\sigma}_k^2 (m-k-1) = \sum_{i=1}^{m-k} \hat{\beta}_{i,k} (F_{i,k} - \hat{f}_k)^2 \quad 1 \le k \le \min(m-2, n-1)$$

If $m > n$ then $\hat{\sigma}_k^2$ are defined for $k = 1,...,n-1$.

Proposition. For $1 \le k \le \min(m-2, n-1)$

$$\hat{\sigma}_k^2 (m-k-1) = \sum_{i=1}^{m-k} \hat{\beta}_{i,k} F_{i,k}^2 - \hat{f}_k^2 \hat{\beta}_k \quad \text{where} \quad \hat{\beta}_k = \sum_{k=1}^{m-k} \hat{\beta}_{i,k}$$

Easy Proof Note $(F_{i,k}-\hat{f}_k)^2 = F_{i,k}^2 - 2F_{i,k}\hat{f}_k + \hat{f}_k^2$ and use the definition $\hat{f}_k = \sum_{i=1}^{m-k} F_{i,k}\hat{b}_{i,k} / \hat{b}_k$ where $\hat{b}_k = \sum_{i=1}^{m-k} \hat{b}_{i,k}$.

Remark. If $m = n$ then we have to find some method of defining $\hat{s}_{n-1}^2$. One possibility would be to use $\hat{s}_{n-2}^2$ or the methods suggested in Mack, 1993-1994-1999. It would be helpful if Schedule P of the U.S. Property and Casualty Annual Statement showed (say) 12 accident years instead of 10 for it would make it easier to compute the standard deviation.

S3. MACK's Formulas and Related Estimators

In this section we complete Mack's proof for the $a = 0, 2$ cases the prove a "summation theorem." We also derive the L-estimators.

Definition. Give any two random variables X and Y the **mean square error** with respect to $D = \{C_{i,k} : i + k \le m+1\}$ is defined as $\text{mse}(X,Y) = E_D(X - Y)^2$. The goals are to estimate the following "mean square errors." (The accident year $q = 2 + m - n$ is the first accident year that is not fully developed.)

(a) $\text{mse}(C_{i,n}, \hat{C}_{i,n}) = E_D(C_{i,n} - \hat{C}_{i,n})^2 \quad (i = q, \ldots, m)$

(b) $\text{mse}(\sum_{i=q}^{m} C_{i,n}, \sum_{i=q}^{m} \hat{C}_{i,n}) = E_D(\sum_{i=q}^{m} C_{i,n} - \sum_{i=q}^{m} \hat{C}_{i,n})^2$

Remark 1. We first split up the above into two parts. (See Mack's papers, Astin 1999 and Astin 1993.) Let i be a fixed accident year.) Thus

$E_D(C_{i,n} - \hat{C}_{i,n})^2 = \text{Var}_D(C_{i,n}) + \{E_D(C_{i,n}) - \hat{C}_{i,n}\}^2$.

(The two terms on the right hand side are computed in Theorem 3A and Theorem 3B.)

Easy Proof. If X is any random variable and $\hat{Y}$ is D-measurable then

$$\text{Var}_D(X) = \text{Var}_D(X - \hat{Y}) = E_D(X - \hat{Y})^2 - E_D^2(X - \hat{Y}) .$$

Remark 2. $\text{mse}(C_{i,n}, \hat{C}_{i,n}) = \text{mse}(R_i, \hat{R}_i)$ for every accident year $i = 1, \ldots, m$.

Easy proof. Let $LTD_i = C_{i,m+1-i}$. We recall

(a) $\hat{R}_i = \hat{C}_{i,n} - LTD_i$

(b) $R_i = C_{i,n} - LTD_i$

Since LTD_i and $\hat{C}_{i,n}$ are D-measurable:

(c) $Var_D(\hat{R}_i) = Var_D(\hat{C}_{i,n})$

(d) $E_D(\hat{R}_i) = E_D(\hat{C}_{i,n}) - LTD_i$

Thus by remark 1: $\text{mse}(C_{i,n}, \hat{C}_{i,n}) = \text{mse}(R_i, \hat{R}_i)$.

Remark 3. Mack (1999 Astin Bulletin) also discusses the following mean square errors:

(c) $\text{mse}(f_k, F_{i,k}) = E_{G_{i,k}}(f_k - F_{i,k})^2$ where $(1 \le i \le m;\ 1 \le k \le n-1)$

(d) $\text{mse}(f_k, \hat{f}_k) = E_{L_k}(f_k - \hat{f}_k)^2$ where $(1 \le k \le n-1)$

Since $E_{G_{i,k}}(F_{i,k}) = f_k$ we have $E_{G_{i,k}}(F_{i,k} - f_k)^2 = \text{Var}_{G_{i,k}}(F_{i,k}) = s_k^2 / b_{i,k}$.

Since $E_{L_k}(\hat{f}_k) = f_k$ we have $E_{L_k}(\hat{f}_k - f_k) = \text{Var}_{L_k}(\hat{f}_k) = s_k^2 / \hat{b}_k$.

Mack (see Astin 1999) defines estimated predictors of the mean square errors as follows:

(c') $\hat{\text{mse}}(f_k, F_{i,k}) = \hat{s}_k^2 / \hat{b}_{i,k}$ where $\hat{b}_{i,k} = w_{i,k} C_{i,k}^{a}$

(d') $\hat{\text{mse}}(f_k, \hat{f}_k) = \hat{s}_k^2 / \hat{b}_k$ where $\hat{b}_k = \sum_{i=1}^{m-k} \hat{b}_{i,k}$

In (c') Mack did not restrict (i,k) to $i + k \le m+1$.

Remark. Let $D = \{C_{i,j} : i + j \le m+1\}$. The D-measurable random variables are unaffected by the operator E_D. In particular:

(a) $E_D(\hat{b}_{i,k}) = \hat{b}_{i,k}$ where $i + k \le m$, $k = 1,....,n-1$

(b) $E_D F_{i,k} = F_{i,k}$ where $i + k \le m$, $k = 1,....,n-1$

(c) $E_D(\hat{f}_k) = \hat{f}_k$ where $k = 1,....,n-1$

(d) $E_D(\hat{s}_k^2) = \hat{s}_k^2$ where $k = 1,....,n-1$

(e) $E_D(\hat{C}_{i,n}) = \hat{C}_{i,n}$ where $1 \le i \le m$

The proofs are easy -- Proof a. Under the hypothesis $i + k \le m$ $\hat{b}_{i,k} = w_{i,k} C_{i,k}^{a}$ is D-measurable, since $C_{i,k} \in D$.

Proof. b. Under the hypothesis $i + k \le m$ both $C_{i,k}$ and $C_{i,k+1}$ are in $D = \{C_{i,j} : i + j \le m+1\}$ and hence are D-measurable and hence the quotient $F_{i,k}$ is D-measurable.

Proof c. By definition $\hat{f}_k = \sum_{i=1}^{m-k} F_{i,k} \hat{b}_{i,k} / \hat{b}_k$ and $\hat{b}_{i,k} = C_{i,k}^{a} w_{i,k}$ and $\hat{b}_k = \sum_{i=1}^{m-k} \hat{b}_{i,k}$ are D-measurable. Also the $F_{i,k}$ are measurable since $i + k \le m+1$. Hence (c) follows from (b).

Proof d. Follows from (b) and (c) which show $\hat{s}_k^2$ is D-measurable.

Proof e. Follows since $\hat{C}_{i,n} = C_{i,p} \hat{f}_p \cdots \hat{f}_{n-1}$ is D-measurable where $p = m + 1 - i$.

Theorem 1A (see Mack Astin 1993, p.215 for all but c2 and d) Let $B_k = \{C_{i,j} : j \le k\}$ and $L_k = \{C_{i,j} : 1 \le i \le m-k;\ j \le k\}$. Under CL1 to CL3 we have:

(a1) $E_{B_k}(F_{i,k}) = E_{L_k}(F_{i,k}) = E_{G_{i,k}}(F_{i,k}) = f_k$ where $(i = 1,...,m)(k = 1,...,n-1)$

(a2) $E(F_{i,k}) = f_k$ where $(i = 1,...,m)$ $(k = 1,...,n-1)$

(b1) $E_{B_k}(\hat{f}_k) = E_{L_k}(\hat{f}_k) = f_k$ where $k = 1,\ldots,n-1$

(b2) $E(\hat{f}_k) = f_k$ where $k = 1,\ldots,n-1$

(c1) $E(\hat{f}_j \ \hat{f}_k) = f_j \ f_k \quad (j < k)$

(c2) $E(\hat{f}_j^2 \ \hat{f}_k) = f_j^2 \ f_k \quad (j < k)$

(d) If $E_{B_k}(\hat{f}_k^2) = g_k^2$ (a constant), then $E(\hat{f}_j^2 \ \hat{f}_k^2) = g_j^2 \ g_k^2 = E(\hat{f}_j^2) E(\hat{f}_k^2)$ for $(j < k)$

Proof a1.

$E_{B_k}(F_{i,k}) = E_{G_{i,k}}(F_{i,k})$ (using CL3 independence of accident rows and $G_{i,k} = B_k \cap A_i$)

$E_{L_k}(F_{i,k}) = E_{G_{i,k}}(F_{i,k})$ (using CL3 independence of accident rows and $G_{i,k} = L_k \cap A_i$)

Finally, $E_{G_{i,k}} F_{i,k} = f_k$ (using CL1).

Proof a2. $E(F_{i,k}) = EE_{G_{i,k}} F_{i,k} = E(f_k) = f_k$ (using property of contingent expectation; (a1); and f_k is a constant).

Proof b1. By definition. $\hat{f}_k = \sum_{i=1}^{m-k} F_{i,k} \boldsymbol{b}_{i,k} / \boldsymbol{b}_k$ so $E_{B_k}(\hat{f}_k) = \sum_{i=1}^{m-k} E_{B_k}(F_{i,k})\hat{\boldsymbol{b}}_{i,k} / \hat{\boldsymbol{b}}_k$ (since $\hat{\boldsymbol{b}}_{i,k} = C_{i,k}^{\boldsymbol{a}} w_{i,k}$ and $\hat{\boldsymbol{b}}_k = \sum_{i=1}^{m-k} \hat{\boldsymbol{b}}_{i,k}$ are B_k and L_k measurable). But $E_{B_k}(F_{i,k}) = f_k$ by (a1). This proves (b1) for B_k The proof for L_k is identical, where B_k is replaced by L_k.

Proof b2. Follows from (b1) since $EE_{B_k} = E$ and $E_{B_k}(\hat{f}_k)$ is a constant.

Proof c1. $E(\hat{f}_j \ \hat{f}_k) = EE_{B_k} \hat{f}_j \ \hat{f}_k \quad (j < k)$ (property of projection operators)

$= E(\hat{f}_j E_{B_k} \hat{f}_k)$ (since $\hat{f}_j$ is B_k measurable for j < k)

$= E(\hat{f}_j \ f_k) = = f_k E(\hat{f}_j)$ (using b1)

$= E(\hat{f}_k) \, E(\hat{f}_j)$ (using b2)

Proof c2 and d. Use the same proof as for (c1).

Theorem 1B (Mack 1993, p215). Let $D = \{C_{i,j} : i + j \le m+1\}$**.** Under the chain ladder assumptions $E(\hat{C}_{i,n}) = E(C_{i,n})$ for every accident year $i = 1,\ldots,m$. In particular:

(a) $E_D(C_{i,n}) = C_{i,p} \ f_p \cdots f_{n-1}$ where $(i = 1,...,m)$ and $p = m+1-i$

(b) $E(C_{i,n}) = E(C_{i,p}) \ f_p \cdots f_{n-1}$ where $(i = 1,...,m)$

(c) $E(\hat{C}_{i,n}) = E(C_{i,p}) f_p \cdots f_{n-1}$ where $(i = 1,...,m)$

Proof of (a). Fix accident year i. Let $G_{i,k} = \{C_{i,j} : 1 \le j \le k\}$ and $D_i = \{C_{i,j} : 1 \le j \le p\}$**.** Then for $k = p+1,\ldots n-1$:

$E_D C_{i,k} = E_{D_i} C_{i,k}$ (using independence and CL3)

$= E_{D_i} E_{G_{i,k-1}} C_{i,k}$ (property of conditional probability since $D_i \subseteq G_{i,n-1}$)

$= E_{D_i}(C_{i,k-1} f_{k-1})$ (using CL1)

$= f_{k-1} E_{D_i}(C_{i,k-1})$ (since f-term is constant).

Now we apply induction and derive:

$E_{D_i} C_{i,n} = f_{n-1} E_{D_i}(C_{i,n-1}) =$

$= f_{m+1-i} \cdots f_{n-2}\, f_{n-1} E_{D_i} C_{i,m+1-i}$ (induction step)

$= f_{m+1-i} \cdots f_{n-2}\, f_{n-1} C_{i,m+1-i}$ since $LTD_i = C_{i,m+1-i}$ is D_i measurable

Proof of (b) Follows by (a) and the following property of conditional expectation: $EE_D = E$.

Proof of (c). Follows since:

(1) $\hat{C}_{i,n} = C_{i,p} \hat{f}_p \cdots \hat{f}_{n-1}$ (by definition)

(2) The term $\hat{f}_k$ depends on accident years $\{1,...,m-k\}$ and

The product $\hat{f}_{p+1} \cdots \hat{f}_{n-1}$ depends on accident years $\{1,...,i-1\}$. Hence $C_{i,p}$ and $\hat{f}_p \cdots \hat{f}_{n-1}$ depend on different accident years and are independent by CL3.

(3) The $\{\hat{f}_p, \ldots, \hat{f}_{n-1}\}$ are uncorrelated; and $E(\hat{f}_k) = f_k$.

Theorem 2 (See Mack, 1994 Cas. Forum pp 151.-153; 1999 Astin pp 361, 363 for (a) and (b).) Let $B_k = \{C_{i,j} : j \le k;\ 1 \le i \le m\}$ Assume $i + k \le m + 1$ and $k = 1,...,n-1$. Under the Chain Ladder assumptions:

(a1) $Var_{B_k} F_{i,k} = \boldsymbol{s}_k^2 / \boldsymbol{b}_{i,k}$ where $\boldsymbol{b}_{i,k} = C_{i,k}^{\boldsymbol{a}} w_{i,k}$

(a2) $Var_{B_k} C_{i,k+1} = C_{i,k}^2 (\boldsymbol{s}_k^2 / \boldsymbol{b}_{i,k})$

(a3) $Var_{B_k}(\hat{f}_k) = \boldsymbol{s}_k^2 / \hat{\boldsymbol{b}}_k$ where $\hat{\boldsymbol{b}}_k = \sum_{i=1}^{m-k} \hat{\boldsymbol{b}}_{i,k}$

(b1) $E_{B_k}(\hat{\boldsymbol{s}}_k^2) = \boldsymbol{s}_k^2$

(b2) $E(\hat{\boldsymbol{s}}_k^2) = \boldsymbol{s}_k^2$

(c1) $E(\hat{f}_j\, \hat{\boldsymbol{s}}_k^2) = f_j\, \boldsymbol{s}_k^2 \ (j \ne k)$

(c2) $E(\hat{f}_j^2\, \hat{\boldsymbol{s}}_k^2) = f_j^2\, \boldsymbol{s}_k^2 \quad (j < k)$

(c3) $E(\hat{\boldsymbol{s}}_k^2 / \hat{\boldsymbol{b}}_k) = E(\hat{\boldsymbol{s}}_k^2) E(1/\hat{\boldsymbol{b}}_k)$

(d1) $E_{B_k}(\hat{f}_k^2) = \boldsymbol{s}_k^2 / \hat{\boldsymbol{b}}_k + f_k^2$

(d2) $E(\hat{f}_k^2) = \boldsymbol{s}_k^2\, E(1/\hat{\boldsymbol{b}}_k) + f_k^2$

Note that in (a1) and (a3) that the results are random variables depending on the claims amounts $\{C_{i,k} : i + k \le m + 1\}$ unless $\boldsymbol{a} = 0$. Note in (a1) to (a3) we can replace B_k by $G_{i,k}$ or $\mathrm{L_k}$ and in (b1) we can replace B_k by L_k.

Proof (a1). By independence (CL3) and by CL2.

$Var_{B_k} F_{i,k} = Var_{G_{i,k}} F_{i,k} = \boldsymbol{s}_k^2 / \boldsymbol{b}_{i,k}$

Proof (a2). Follows from (a1).

Proof (a3). By definition: $\hat{f}_k = (1/\hat{b}_k)\sum_{i=1}^{m-k} F_{i,k}\hat{b}_{i,k}$. Since the $\hat{b}_{i,k}$ are B_k measurable we have

$$Var_{B_k}(\hat{f}_k) = (1/\hat{b}_k^2)\sum_{i=1}^{m-k} \hat{b}_{i,k}^2\, Var_{B_k}(F_{i,k}).$$

Using (a1) completes the proof.

Proof (b1). By definition:

$$(m-k-1)\hat{s}_k^2 = \sum_{i=1}^{m-k} \hat{b}_{i,k} F_{i,k}^2 - \hat{b}_k(\hat{f}_k^2)$$

Note that $b_{i,k}$ and b_k are B_k measurable. so by (a)

$$(m-k-1)E_{B_k}(\hat{s}_k^2) = \sum_{i=1}^{m-k} \hat{b}_{i,k} E_{B_k}(F_{i,k}^2) - \hat{b}_k E_{B_k}(\hat{f}_k^2)$$

Recall $E(X^2) = Var(X) + E^2(X)$ for any integrable random variable X. Also

$Var_{B_k} F_{i,k} = s^2/\hat{b}_{i,k}$ and $E_{B_k}(F_{i,k}) = f_k$.

$Var_{B_k}(\hat{f}_k) = s_k^2/\hat{b}_k$ and $E_{B_k}(\hat{f}_k) = f_k$.

Thus $(m-k-1)E_{B_k}(\hat{s}_k^2) = \sum_{i=1}^{m-k} \hat{b}_{i,k}\{\frac{s_k^2}{\hat{b}_{i,k}} + f_k^2\} - \hat{b}_k\{\frac{s_k^2}{\hat{b}_k} + f_k^2\} =$

$= (m-k)s_k^2 - s_k^2 = (m-k-1)s_k^2$.

Proof (b2) Follows from (b1) since $E(X) = EE_{B_k}(X)$ for any random variable X.

Proof (c1) Follows as in theorem 1A using the results above. Thus for j < k:

$E_{B_k}(\hat{f}_j\, \hat{s}_k^2) = \hat{f}_j\, E_{B_k}(\hat{s}_k^2)$ since $\hat{f}_j$ is B_k measurable and

$E_{B_k}(\hat{s}_j^2\, \hat{f}_k) = \hat{s}_j^2\, E_{B_k}(\hat{f}_k)$ since $\hat{s}_j$ is B_k measurable.

Proof (c2) Similar to the above.

Proof (c3) $EE_{B_k}(\hat{s}_k^2/\hat{b}_k) = E(1/\hat{b}_k E_{B_k}(\hat{s}_k^2))$ **(since $1/\hat{b}_k$ is B_k measurable)**

$= E(1/\hat{b}_k)s_k^2$ **(since $E_{B_k}(\hat{s}_k^2)$ is the constant s_k^2)**

$= E(1/\hat{b}_k)E(\hat{s}_k^2)$ **(by b2).**

Proof (d1) Use $E_{B_k}(\hat{f}_k^2) = Var_{B_k}(\hat{f}_k) + E_{B_k}^2(\hat{f}_k) = s_k^2/\hat{b}_k + f_k^2$.

Proof (d2). Follows from (d1).

Theorem 3A. (cf. Mack Astin 1999 for the statements of (a) and (b) -- but we have modified a0 and b0) Let i be a fixed accident year and let $p = m+1-i$ so that $C_{i,p}$ is the losses to date for accident year i. For $a = 0$ let $g_{i,k}^2 = f_k^2 + s_k^2/w_{i,k}$.

Then by the chain ladder assumptions we have the following for $k = p+1,\ldots,n-1$.

(a0) $\alpha = 0$: $\mathrm{Var}_D(C_{i,k+1}) = \{\sigma_k^2 / w_{i,k}\} f_{k-1}^2 \cdots f_p^2 C_{i,p}^2 + g_{i,k}^2 \mathrm{Var}_D(C_{i,k})$

(a1) $\alpha = 1$: $\mathrm{Var}_D(C_{i,k+1}) = \{\sigma_k^2 / w_{i,k}\} f_{k-1} \cdots f_p C_{i,p} + f_k^2 \mathrm{Var}_D(C_{i,k})$

(a2) $\alpha = 2$: $\mathrm{Var}_D(C_{i,k+1}) = \{\sigma_k^2 / w_{i,k}\} + f_k^2 \mathrm{Var}_D(C_{i,k})$

We have the following formulas for the variance of $C_{i,n}$.

(b0) $\alpha = 0$: $\mathrm{Var}_D(C_{i,n}) = C_{i,p}^2 \sum_{k=p}^{n-1} f_p^2 \cdots f_{k-1}^2 \{\sigma_k^2 / w_{i,k}\} g_{i,k+1}^2 \cdots g_{i,n-1}^2$

(b1) $\alpha = 1$: $\mathrm{Var}_D(C_{i,n}) = C_{i,p} \sum_{k=p}^{n-1} f_p \cdots f_{k-1} \{\sigma_k^2 / w_{i,k}\} f_{k+1}^2 \cdots f_{n-1}^2$

(b2) $\alpha = 2$: $\mathrm{Var}_D(C_{i,n}) = \sum_{k=p}^{n-1} \{\sigma_k^2 / w_{i,k}\} f_{k+1}^2 \cdots f_{n-1}^2$

Proof -- notation

Let $D = \{C_{i,1}, ..., C_{i,p}\}$ be the known values for accident year i and let $G = \{C_{i,1}, ..., C_{i,k}\}$ where $k = p+1, \ldots n-1$. Note that $D \subseteq G$. Also note for any random variable $E_D(X^2) = E_D^2(X) + \mathrm{Var}_D(X)$.

Proof (a0)

$$\mathrm{Var}_D(C_{i,k+1}) = E_D \mathrm{Var}_G(C_{i,k+1}) + \mathrm{Var}_D E_G(C_{i,k+1}) =$$
$$= \{\sigma_k^2 / w_{i,k}\} E_D(C_{i,k}^2) + f_k^2 \mathrm{Var}_D(C_{i,k}) =$$
$$= \{\sigma_k^2 / w_{i,k}\} E_D^2(C_{i,k}) + (\sigma_k^2 / w_{i,k} + f_k^2) \mathrm{Var}_D(C_{i,k}) =$$
$$= \{\sigma_k^2 / w_{i,k}\} f_{k-1}^2 \cdots f_p^2 C_{i,p}^2 + g_{i,k}^2 \mathrm{Var}_D(C_{i,k}) .$$

Proof (a1)

$$\mathrm{Var}_D(C_{i,k+1}) = E_D \mathrm{Var}_G(C_{i,k+1}) + \mathrm{Var}_D E_G(C_{i,k+1}) =$$
$$= \{\sigma_k^2 / w_{i,k}\} E_D(C_{i,k}) + f_k^2 \mathrm{Var}_D(C_{i,k}) =$$
$$= \{\sigma_k^2 / w_{i,k}\} f_{k-1} \cdots f_p C_{i,p} + f_k^2 \mathrm{Var}_D(C_{i,k})$$

Proof (a2)

$$\mathrm{Var}_D(C_{i,k+1}) = E_D \mathrm{Var}_G(C_{i,k+1}) + \mathrm{Var}_D E_G(C_{i,k+1}) =$$
$$= \{\sigma_k^2 / w_{i,k}\} E_D(1) + f_k^2 \mathrm{Var}_D(C_{i,k}) = \{\sigma_k^2 / w_{i,k}\} + f_k^2 \mathrm{Var}_D(C_{i,k})$$

Proof (b0)-(b1)-(b2). Follows by induction.. Note that $C_{i,p}$ is D-measurable so $E_D(C_{i,p}) = C_{i,p}$ and $\mathrm{Var}_D(C_{i,p}) = 0$.

Remark. In Astin 1999 Mack gives formulas for the $\alpha = 0$ case, but he used $\hat{f}_k^2$ in place of $\hat{g}_{i,k}^2 = \hat{f}_k^2 + \hat{\sigma}_k^2 / w_{i,k}$. This was probably deliberate.

Theorem 3B. (see Mack 1993, 1994 for most of the proof; we added b2, c3 and d) Let i be a fixed accident year and let $p = m + 1 - i$. Let $V_i = \{E_D(C_{i,n}) - \hat{C}_{i,n}\}^2$ and let

$$\hat{V}_i = C_{i,p}^2 \sum_{k=p}^{n-1} (\hat{f}_p^2 \cdots \hat{f}_{k-1}^2 / \hat{b}_k)(s_k^2) f_{k+1}^2 \cdots f_{n-1}^2 .$$

Under the Chain Ladder assumptions $E(V_i) = E(\hat{V}_i)$.

Proof. We prove the following steps:

(a1) $V_i = \{E_D(C_{i,n}) - \hat{C}_{i,n}\}^2 = C_{i,p}^2 (f_p^2 \cdots f_{n-1}^2 - \hat{f}_p^2 \cdots \hat{f}_{n-1}^2)$

$$= C_{i,p}^2 \left(\sum_{k=p}^{n-1} S_k^2 + \sum_{j<k} S_j S_k \right)$$

where $S_k = \hat{f}_p \cdots \hat{f}_{k-1}(f_k - \hat{f}_k)(f_{k+1} \cdots f_{n-1})$ and $p = m + 1 - i$.

(a2) $C_{i,p}$ and $\hat{f}_p^2 \cdots \hat{f}_{n-1}^2$ are independent. (where $p = m + 1 - i$)

(b1) $E_{B_k}(S_j S_k) = 0$ for $j < k$

(b2) $E(S_j S_k) = 0$ for $j < k$

(c1) $E_{B_k}(f_k - \hat{f}_k) = s_k^2 / \hat{b}_k$ where $\hat{b}_k = \sum_{g=1}^{m-k} \hat{b}_{i,k}$

(c2) $E_{B_k}(S_k^2) = \hat{f}_p^2 \cdots \hat{f}_{k-1}^2 (s_k^2 / \hat{b}_k) f_{k+1}^2 \cdots f_{n-1}^2$ where $k = p, \ldots, n-1$

(c3) $E(S_k^2) = E(\hat{f}_p^2 \cdots \hat{f}_{k-1}^2 / \hat{b}_k)\ (s_k^2) f_{k+1}^2 \cdots f_{n-1}^2$ where $k = p, \ldots, n-1$

(d) $E(V_i) = E(\hat{V}_i)$.

Proof (a1). First equality follows by theorem 1B. The second follows from a device introduced by Mack. Let

$$S_p = (f_p - \hat{f}_p) f_{p+1} \cdots f_{n-1}$$

$$S_{p+1} = \hat{f}_p(f_{p+1} - \hat{f}_{p+1}) f_{p+2} \cdots f_{n-1}$$

$$S_{p+2} = \hat{f}_p \hat{f}_{p+1}(f_{p+2} - \hat{f}_{p+2}) f_{p+3} \cdots f_{n-1} \text{, etc.}$$

Then the S_k equals $(f_p^2 \cdots f_{n-1}^2 - \hat{f}_p^2 \cdots \hat{f}_{n-1}^2)$.

Proof (a2) The term $C_{i,p}$ depends on accident year i and $\hat{f}_p$ and the product $\hat{f}_p^2 \cdots \hat{f}_{n-1}^2$ depends on accident years $\{1,...,i-1\}$ and are independent by CL3.

Proof (b1) Apply the operator E_{B_k} to $S_j S_k$. Since $(f_{k+1} \cdots f_{n-1})$ are constants and S_j and $\hat{f}_p \cdots \hat{f}_{k-1}$ are B_k measurable they factor out, and:

$$E_{B_k}(S_j S_k) = S_j \hat{f}_p \cdots \hat{f}_{k-1}(f_{k+1} \cdots f_{n-1}) E_{B_k}(f_k - \hat{f}_k)$$

Now the above equals zero since $E_{B_k}(f_k - \hat{f}_k) = 0$.

Proof (b2) Follows from (b1) since $EE_{B_k}(X) = E(X)$ for every random variable X.

Proof (c1) $E_{B_k}(\hat{f}_k) = f_k$ **so** $E_{B_k}(f_k - \hat{f}_k) = Var_{B_k}(\hat{f}_k) = \mathbf{s}_k^2 / \hat{\mathbf{b}}_k$ **.**

Proof (c2). Follows since $\hat{f}_p^2 \cdots \hat{f}_{k-1}^2$ and $\hat{\mathbf{b}}_k = \sum_{g=1}^{m-k} w_{gk} C_{gk}^{\mathbf{a}}$ are B_k measurable for k ³ p.

Proof (c3) Follows from (c1)

Proof (d) $E(V_i) = E(C_{i,p}^2)\{\sum_{k=p}^{n-1} E(S_k^2) + \sum_{j<k} E(S_j S_k)\}$ **(using a2, independence)**

$= E(C_{i,p}^2) \sum_{k=p}^{n-1} E(\hat{f}_p^2 \cdots \hat{f}_{k-1}^2 / \hat{\mathbf{b}}_k)\ (\mathbf{s}_k^2) f_{k+1}^2 \cdots f_{J-1}^2$ **(using (b) and (c))**

Note. In the $\mathbf{a} = 0$ case $\mathbf{b}_k$ is constant and the squares $\{\hat{f}_j^2 : j = 1, \ldots m-1\}$ are uncorrelated and $E(\hat{f}_k^2) = E^2(\hat{f}_k) + \mathrm{Var}(\hat{f}_k) = f_k^2 + \mathbf{s}_k^2 / \hat{\mathbf{b}}_k = g_k^2$.

THEOREM 4 A. (Cf. Mack Cas Forum 1994 p 153; we have added some details to the proof). Let q = 2+m-n be the first accident year that is not fully developed. Under the chain ladder $\mathrm{mse}(R, \hat{R})$ equals

$$= \sum_{i=q}^{m} \mathrm{mse}(C_{i,n}, \hat{C}_{i,n}) + 2 \sum_{i<g} (C_{i,n} - \hat{C}_{i,n})(C_{g,n} - \hat{C}_{g,n})$$

Proof (cf. Mack 1993 page 200 and cf. Greg Taylor, 2000)

$$\mathrm{mse}(R, \hat{R}) = E_D(\sum_{i=q}^{m} E(C_{i,n}) - \sum_{i=q}^{m} E(\hat{C}_{i,n}))^2 \quad \textbf{(by definition)}$$

$$= \mathrm{Var}_D(\sum_{i=q}^{m} C_{i,n}) + E_D(\sum_{i=q}^{m} C_{i,n} - \sum_{i=q}^{m} \hat{C}_{i,n})^2 \quad \textbf{(by property of variance; detail 1)}$$

$$= \sum_{i=q}^{m} \mathrm{Var}(C_{i,n}) + E_D(\sum_{i=q}^{m} C_{i,n} - \sum_{i=q}^{m} \hat{C}_{i,n})^2 \quad \textbf{(by detail 2 and the summation theorem)}$$

$$= \sum_{i=q}^{m} \mathrm{Var}(C_{i,n}) + E_D \sum_{i=q}^{\mathrm{m}} (C_{i,n} - \hat{C}_{i,n})^2 + 2E_D \sum_{i<g} (C_{i,n} - \hat{C}_{i,n})(C_{g,n} - \hat{C}_{g,n})$$

$$= \sum_{i=1}^{m} \mathrm{mse}(C_{i,n}, \hat{C}_{i,n}) + 2E_D \sum_{i<g} (C_{i,n} - \hat{C}_{i,n})(C_{g,n} - \hat{C}_{g,n}) \quad \textbf{(by theorem 3)}$$

$$= \sum_{i=q}^{m} \mathrm{mse}(C_{i,n}, \hat{C}_{i,n}) + 2E_D \sum_{i=q}^{m-1} (C_{i,n} - \hat{C}_{i,n}) \sum_{g=i+1}^{m} (C_{g,n} - \hat{C}_{g,n})$$

Detail 1. This was shown earlier. In general if D is any set of integrable random variables and $\hat{Y}$ is D-measurable, then: $E_D(X - \hat{Y})^2 - \{E_D(X) - \hat{Y}\}^2 = \mathrm{Var}_D(X - \hat{Y}) = \mathrm{Var}_D(X)$

Detail 2. If $E_D(XY) = E_D(X)E_D(Y)$, then $\mathrm{Var}_D(X+Y) = \mathrm{Var}_D(X) + \mathrm{Var}_D(Y)$. But if X, Y are uncorrelated or independent that does not mean that the conditional expectations are uncorrelated or uncorrelated, see Jordan Stoyanov, Counterexamples in Probability, 7.1.3 page 55.. We need the following “summation theorem”:

Summation Theorem.

$$E_D(C_{g,k}\,C_{i,k}) = E_{D_g \vee D_i}(C_{g,k}\,C_{i,k}) = E_{D_g}(C_{g,k})E_{D_i}(C_{i,k}) = E_D(C_{g,k})E_D(C_{i,k})$$

where $D_g \vee D_i$ is the $s-$algebra generated by the functions in $D_g \cup D_i$.

Proof. We will We will rephrase the above in terms of $s-$algebras. Let A_g and A_i be the $s-$algebras D_g and D_i respectively. generated by

The first and last equation of (a) follows from CL3 (independence of the rows) and properties of contingent expectation. The second equation of (a) follows by showing that the third term has the unique properties of the second term -- these are properties (1) and (2.3) below.

We will now prove the following:

(1) $E_{D_g}(C_{g,k})E_{D_i}(C_{i,k})$ is $D_g \vee D_i$ measurable.

(2.1) $V = (C_{g,n}\,C_{i,n}) - E_{D_g}(C_{g,n})E_{D_i}(C_{i,n})$ is perpendicular to all products YZ where Y is D_g measurable and Z is D_i measurable.

(2.2) V is perpendicular to $s-$algebra generated by the union of D_g and D_i.

(2.3) V is a measurable function with respect to $D_g \cup D_i$

Proof (1). Property (1) is trivial since $E_{D_g}(C_{g,n})$ is D_g measurable and $E_{D_i}(C_{i,n})$ is D_i measurable and hence the product is measurable for any $s-$algebra which contains $D_g \cup D_i$

Proof (2.1) Note that $\{C_{g,n},\, E_{D_g}(C_{g,n}),\, Y\}$ are D_g measurable and hence pairwise independent of the D_i measurable functions $\{C_{i,n},\, E_{D_i}(C_{i,n}),\, Z\}$. Hence

$$E(YVZ) = E(Y\,C_{g,n}\,C_{i,n}\,Z) - E(Y\,E_{D_g}(C_{g,n})E_{D_i}(C_{i,n})\,Z) =$$

$$= E(Y\,C_{g,n})\,E(C_{i,n}\,Z) - E\{Y\,E_{D_g}(C_{g,n})\}\,E\{E_{D_i}(C_{i,n})\,Z\}$$ (pairwise independence).

But $E(Y\,C_{g,n}) = E\{Y\,E_{D_g}(C_{g,n})\}$ and $E(C_{i,n}\,Z) = E\{E_{D_i}(C_{i,n})\,Z\}$

by definition of the projection operators. Therefore V ^ ZY.

Proof (2.2). Let A_g and A_i be the $s-$algebras generated by the functions in D_g and D_i respectively. Item (2.1) implies that V is perpendicular to all the indicators in $A_g \cup A_i$. Perpendicularity preserves unions, intersections and limits. Hence V is perpendicular to the $s-$algebra generated by $A_g \cup A_i$

Proof (2.3). This follows since the minimum s algebra which contains $D_g \cup D_i$ contains V.

Note. For a proof similar to the above see J.L. Doob Measure theory, page 186, proof of (k). and page 23 definition of independent $s-$algebras

The notation $D_g \vee D_i$ is defined in Billingsley, Probability and Measure (3rd ed) at 455.

Theorem 4B. (See Mack for most of the proof**;** we added some details to show $E(Z) = E(\hat{Z})$**)**
Assume the chain ladder hypothesis; with q = 2+m-n the first accident year that is not fully developed.

Let $Z = 2\sum_{i=q}^{m-1} (E_D C_{i,n} - \hat{C}_{i,n}) \sum_{g=i+1}^{m} (E_D C_{g,n} - \hat{C}_{g,n})$.

Let $\hat{Z} = 2\sum_{i=q}^{m-1} C_{i,m+1-i} \sum_{k=m+1-i}^{n-1} \hat{f}_{m+1-i}^2 \cdots \hat{f}_{k-1}^2 (s_k^2 / \hat{b}_k) f_{k+1}^2 \cdots f_{n-1}^2 \sum_{g=i+1}^{m} \hat{C}_{g,m+1-i}$.

Then $E(Z) = E(\hat{Z})$**.**

Proof. For fixed i and g let us examine $X0 = (E_D C_{i,n} - \hat{C}_{i,n})(E_D C_{g,n} - \hat{C}_{g,n})$.

We find for $i < g$

(1) $E_D C_{g,n} = C_{g,m+1-g}(f_{m+1-g} \cdots f_{m-i})(f_{m+1-i} \cdots f_{n-1})$

(2) $\hat{C}_{g,n} = C_{g,m+1-g}(\hat{f}_{m+1-g} \cdots \hat{f}_{m-i})(\hat{f}_{m+1-i} \cdots \hat{f}_{n-1})$

(3) $E_D C_{i,n} = C_{i,m+1-i}(f_{m+1-i} \cdots f_{n-1})$

(4) $\hat{C}_{i,n} = C_{i,m+1-i}(\hat{f}_{m+1-i} \cdots \hat{f}_{n-1})$**.**

Let $h_1 = (f_{m+1-g} \cdots f_{m-i})$; $h_2 = (f_{m+1-i} \cdots f_{n-1})$

$\hat{h}_1 = (\hat{f}_{m+1-g} \cdots \hat{f}_{m-i})$; $\hat{h}_2 = (\hat{f}_{m+1-i} \cdots \hat{f}_{n-1})$**.**

Then $(E_D C_{g,n} - \hat{C}_{g,n}) = (h_1 h_2 - \hat{h}_1 \hat{h}_2) C_{g,m+1-g}$ and $(E_D C_{i,n} - \hat{C}_{i,n}) = (h_2 - \hat{h}_2) C_{i,m+1-i}$

Thus $X0 = \{h_1 h_2 - \hat{h}_1 \hat{h}_2\} C_{g,m+1-g} \{h_2 - \hat{h}_2\} C_{i,m+1-i} =$

$= \hat{h}_1 \{h_2 - \hat{h}_2\} C_{g,m+1-g} \{h_2 - \hat{h}_2\} C_{i,m+1-i} + (h_1 - \hat{h}_1) h_2 C_{g,m+1-g} \{h_2 - \hat{h}_2\} C_{i,m+1-i}$**.**

The expectation of the second term is zero, for apply $E E_{B_{n-1}} \cdots E_{B_{m+1-i}}$ to the second term and note that $E_{B_{m+1-i}} \cdots E_{B_{n-1}} (h_2 - \hat{h}_2) = 0$; see detail 1.

Let $X1$ be the first term: $X1 = \hat{h}_1 C_{g,m+1-g} \{h_2 - \hat{h}_2\}^2 C_{i,m+1-i}$**.** Since $\hat{h}_1 C_{g,m+1-g} = \hat{C}_{g,m+1-i}$ we have

$X1 = \hat{C}_{g,m+1-i} \{h_2 - \hat{h}_2\}^2 C_{i,m+1-i}$.

Note that the three factors of $X1$ depend on different accident years --- $g, \{1,...,i-1\}, i$ -- and hence are independent.

We proceed as in the proof of theorem 3 to evaluate the expectation of $\{h_2 - \hat{h}_2\}^2$. We find:

$$\{h_2 - \hat{h}_2\}^2 = (f_{m+1-i} \cdots f_{n-1})^2 - (\hat{f}_{m+1-i} \cdots \hat{f}_{n-1})^2 =$$

$$= (\sum_{k=m+1-i}^{n-1} S_k^2 + \sum_{j<k} S_j S_k)\text{, where } S_k = \hat{f}_p \cdots \hat{f}_{k-1}(f_k - \hat{f}_k)(f_{k+1} \cdots f_{n-1})$$

Now $E_{B_k}(S_j S_k) = 0 = E(S_j S_k)$ and

$E_{B_k}(f_k - \hat{f}_k)^2 = s_k^2 / \hat{b}_k$ and $E(f_k - \hat{f}_k)^2 = EE_{B_k}(f_k - \hat{f}_k)^2 = E(s_k^2 / \hat{b}_k)$

Thus

$E(S_k^2) = \hat{f}_{m+1-i}^2 \cdots \hat{f}_{k-1}^2 (s_k^2 / \hat{b}_k) f_{k+1}^2 \cdots f_{n-1}^2$ for $k = m+1-i, \ldots, n-1$.

Thus $E(X1) = E(\hat{C}_{g,m+1-i}) E(C_{i,m+1-i}) E\{(\hat{h}_1 - \hat{h}_2)^2\} =$

$$= E(\hat{C}_{g,m+1-i}) E(C_{i,m+1-i}) \sum_{k=m+1-i}^{n-1} E(\hat{f}_{m+1-i}^2 \cdots \hat{f}_{k-1}^2 / \hat{b}_k)(s_k^2) f_{k+1}^2 \cdots f_{n-1}^2 .$$

By taking sums we find $E(Z) = E(\hat{Z})$ as stated in the theorem.

Detail 1. Let $X = (h_1 - \hat{h}_1) h_2 C_{g,m+1-g} C_{i,m+1-i}$ We will show $E(X\{h_2 - \hat{h}_2\}) = 0.$

The proof uses the following steps:

(1) $E_{B_{m+1-i}} \cdots E_{B_{n-1}}(\hat{h}_2) = h_2 = (f_{m+1-i} \cdots f_{n-1})$

(2) $E_{B_{m+1-i}} \cdots E_{B_{n-1}}(h_2 - \hat{h}_2) = 0$

(3) $E_{B_{m+1-i}} \cdots E_{B_{n-1}}(X\{h_2 - \hat{h}_2\}) = 0$

(4) $E(X\{h_2 - \hat{h}_2\}) = 0.$

Proof (1). Recall $\hat{h}_2 = (\hat{f}_{m+1-i} \cdots \hat{f}_{n-1})$. Since $(\hat{f}_{m+1-i} \cdots \hat{f}_{n-2})$ is B_{n-1} measurable we find

$$E_{B_{n-1}}(\hat{f}_{m+1-i} \cdots \hat{f}_{n-1}) = (\hat{f}_{m+1-i} \cdots \hat{f}_{n-2}) E_{B_{n-1}}(\hat{f}_{n-1}) = (\hat{f}_{m+1-i} \cdots \hat{f}_{n-2}) f_{n-1}.$$

Then (1) follows by induction.

Proof (2) Follows by (1).

Proof (3) Follows since X is measurable with respect to $E_{B_{m+1-i}}; \cdots ; E_{B_{n-1}}$ so

$$E_{B_{m+1-i}} \cdots E_{B_{n-1}}(X\{h_2 - \hat{h}_2\}) = X\, E_{B_{m+1-i}} \cdots E_{B_{n-1}}\{h_2 - \hat{h}_2\} = 0.$$

Proof (4) Follows from (4) since $EE_{B_{m+1-i}} \cdots E_{B_{n-1}}(Y) = E(Y)$ for every random variable Y.

S3.2. Formulas for the Mean Square Errors

In the above theorems we computed $U_i = Var_D(C_{i,n})$, $V_i = \{E_D(C_{i,n}) - \hat{C}_{i,n}\}^2$, $Y = \sum_{i=1}^{m} mse(C_{i,n}, \hat{C}_{i,n})$ and $Z = 2\sum_{i=q}^{m-1}(C_{i,n} - \hat{C}_{i,n})\sum_{g=i+1}^{m}(C_{g,n} - \hat{C}_{g,n})$. We also computed some unbiased predictors $\hat{V}_i, \hat{Y}, \hat{Z}$ as shown below.

Summary of the formulas from theorems 3 and 4.

(a0) $a = 0$: $U_i = \mathrm{Var}_D\, C_{i,n} = C_{i,p}^2 \sum_{k=p}^{n-1} f_p^2 \cdots f_{k-1}^2 \{s_k^2 / w_{i,k}\} g_{i,k+1}^2 \cdots g_{i,n-1}^2$

(a1) $a = 1$: $U_i = \mathrm{Var}_D\, C_{i,n} = C_{i,p} \sum_{k=p}^{n-1} f_p \cdots f_{k-1} \{s_k^2 / w_{i,k}\}\ f_{k+1}^2 \cdots f_{n-1}^2$

(a2) $a = 2$: $U_i = \mathrm{Var}_D\, C_{i,n} = \sum_{k=p}^{n-1} \{s_k^2 / w_{i,k}\} f_{k+1}^2 \cdots f_{n-1}^2$

(b) $V = \{E_D(C_{i,n}) - \hat{C}_{i,n}\}^2 \approx \hat{V}_i = C_{i,p}^2 \sum_{k=p}^{n-1} \hat{f}_p^2 \cdots \hat{f}_{k-1}^2 (s_k^2 / \hat{b}_k) f_{k+1}^2 \cdots f_{n-1}^2$

(c) $Y = \sum_{i=q}^{m} mse(C_{i,n}, \hat{C}_{i,n}) \approx \hat{Y} = \sum_{i=q}^{m} U_i + \hat{V}_i$

(d) $Z = 2\sum_{i=q}^{m-1}(E_D C_{i,n} - \hat{C}_{i,n}) \sum_{g=i+1}^{m}(E_D C_{g,n} - \hat{C}_{g,n}) \approx \hat{Z} =$

$$= 2\sum_{i=1}^{\mathrm{m}} C_{i,m+1-i} \sum_{k=m+1-i}^{n-1} \hat{f}_{\mathrm{m}+1-i}^2 \cdots \hat{f}_{k-1}^2 (s_k^2 / \hat{b}_k) f_{k+1}^2 \cdots f_{n-1}^2 \sum_{g=i+1}^{m} \hat{C}_{g,m+1-i}$$

With the predictors $\hat{V}_i, \hat{Y}, \hat{Z}$ we are not done because they involve unknown parameters like f_k, f_k^2, s_k^2. Therefore to compute predictors of the mean square error we need to replace these parameters by estimators. We give formulas for (1) Mack predictors and in (2) for what we call the "L-predictors."

(1) Mack's Predictors. These predictors are based on replacing f_k, s_k^2, f_k^2 by $\hat{f}_k, \hat{s}_k^2, \hat{f}_k^2$ and $g_{i,k}^2$ by $\hat{f}_k^2$. Mack's predictors have an advantage in that the mean square error can be expressed in a simple formula in terms of the estimated claims.

(2) L-Predictors. These predictors are based on replacing the parameters $f_k, s_k^2, g_{i,k}^2, f_k^2, E(\hat{s}_k^2 / \hat{b}_k)$ by their unbiased estimators. In the $a = 0$ case the L-predictors produce an unbiased estimator of the mean square error. We call them L-Predictors because they are based on a property of maximum likelihood estimators.

The substitutions are shown in the chart below.

	Parameter	L-Estimator	Mack's Estimator
1	f_k	$\hat{f}_k$	$\hat{f}_k$
2	s_k^2	$\hat{s}_k^2$	$\hat{s}_k^2$
3	$g_{i,k}^2 = f_k^2 + s_k^2 / w_{i,k}$	$\hat{g}_{i,k}^2 = \hat{h}_k^2 + \hat{s}_k^2 / w_{i,k}$	$\hat{f}_k^2$
4	f_k^2	$\hat{h}_k^2 = \hat{f}_k^2 - \hat{s}_k^2 / \hat{b}_k$	$\hat{f}_k^2$
5	$E(s_k^2 / \hat{b}_k)$	$\hat{s}_k^2 / \hat{b}_k$	$\hat{s}_k^2 / \hat{b}_k$

One cannot "prove" that either the L-predictors or the Mack predictors are "correct"-- they are estimates. If $a = 0$ the L-predictor formulas are unbiased but we have not investigated the bias in the other cases.

Mack Predictors:

a0. $a = 0 \quad \mathrm{Var}_D\, C_{i,n} = U_i \approx \hat{C}_{i,n}^2 \sum_{k=p}^{n-1} \hat{s}_k^2 / \{ \hat{f}_k^2 w_{i,k} \}$

a1. $a = 1 \quad \mathrm{Var}_D\, C_{i,n} = U_i \approx \hat{C}_{i,n}^2 \sum_{k=p}^{n-1} \hat{s}_k^2 / \{ \hat{f}_k^2 \hat{C}_{i,k} w_{i,k} \}$

a2. $a = 2 \quad \mathrm{Var}_D\, C_{i,n} = U_i \approx \hat{C}_{i,n}^2 \sum_{k=p}^{n-1} \hat{s}_k^2 / \{ \hat{f}_k^2 \hat{C}_{i,k}^2 w_{i,k} \}$

b. $\{ E_D(C_{i,n}) - \hat{C}_{i,n} \}^2 \approx \hat{V}_i \approx \hat{C}_{i,n}^2 \sum_{k=p}^{n-1} \hat{s}_k^2 / \{ \hat{f}_k^2 \hat{b}_k \}$

c. $\sum_{i=q}^{m} mse(C_{i,n}, \hat{C}_{i,n}) \approx \hat{Y} = \sum_{i=q}^{m} U_i + \hat{V}_i$

d. $Z = 2 \sum_{i=q}^{m-1} (C_{i,n} - \hat{C}_{i,n}) \sum_{g=i+1}^{m} (C_{g,n} - \hat{C}_{g,n}) \approx \hat{Z}_i \approx \sum_{i=q}^{m-1} \hat{C}_{i,n}\ Fact2_i\ Fact3_i$

where $Fact2_i = \sum_{g=i+1}^{m} \hat{C}_{g,n}$ **and** $Fact3_i = \sum_{k=m+1-i}^{n-1} \frac{2 \hat{s}_k^2}{\hat{b}_k\ \hat{f}_k^2}$

where $\hat{b}_k = \sum_{i=1}^{m-k} \hat{b}_{i,k}$ and $\hat{b}_{i,k} = w_{i,k} C_{i,k}^{a}$ and $a \in \{0,1,2\}$

Derivation of Mack Formulas from theorem 3 and 4. We use $\hat{f}_p \cdots \hat{f}_k = \hat{C}_{i,k} / \hat{C}_{i,p}$ and $\hat{f}_p \cdots \hat{f}_{n-1} = \hat{C}_{i,n} / \hat{C}_{i,p}$. We give more details for $\hat{Z}_i$ formula. Let $Fact1A_i = C_{i,m+i-i} = C_{i,p}$; $Fact2A_i = \sum_{g=i+1}^{m} \hat{C}_{g,m+1-i}$ and let $Fact3A_i = 2 \sum_{k=m+1-i}^{n-1} \hat{f}_{m+1-i}^2 \cdots \hat{f}_{k-1}^2 (s_k^2 / \hat{b}_k) f_{k+1}^2 \cdots f_{n-1}^2$.

Written for and presented at 8th GCA, Mumbai 10-11 March, 2006

Then $Fact1A_i\,(\hat{C}_{i,n}/C_{i,p}) = \hat{C}_{i,n}$ and $Fact2A_i\,(\hat{C}_{i,n}/C_{i,p}) = Fact2_i$ while

$$Fact3A_i = (\hat{f}_p^2 \cdots \hat{f}_{n-1}^2) \sum_{k=m+1-i}^{n-1} \frac{2\hat{s}_k^2}{\hat{b}_k\,\hat{f}_k^2} = FACT3_i\,(\hat{C}_{i,n}^2/C_{i,p}^2).$$

L Predictors

a0. $a = 0$ $\mathrm{Var}_D(C_{i,n}) = U_i \approx C_{i,p}^2 \sum_{k=p}^{n-1} \hat{h}_p^2 \cdots \hat{h}_{k-1}^2 \{\hat{s}_k^2 / w_{i,k}\} \hat{g}_{i,k+1}^2 \cdots \hat{g}_{i,n-1}^2$

a1. $a = 1$ $\mathrm{Var}_D(C_{i,n}) = U_i \approx C_{i,p} \sum_{k=p}^{n-1} \hat{f}_p \cdots \hat{f}_{k-1} \{\hat{s}_k^2 / w_{i,k}\} \hat{h}_{k+1}^2 \cdots \hat{h}_{n-1}^2$

a2. $a = 2$ $\mathrm{Var}_D(C_{i,n}) = U_i \approx \sum_{k=p}^{n-1} \{\hat{s}_k^2 / w_{i,k}\} \hat{h}_{k+1}^2 \cdots \hat{h}_{n-1}^2$

b. $\{E_D(C_{i,n}) - \hat{C}_{i,n}\}^2 \approx \hat{V}_i \approx \sum_{k=p}^{n-1} \hat{f}_p^2 \cdots \hat{f}_{k-1}^2 (\hat{s}_k^2 / \hat{b}_k) \hat{h}_{k+1}^2 \cdots \hat{h}_{n-1}^2$

c. $\sum_{i=q}^{m} mse(C_{i,n}, \hat{C}_{i,n}) \approx \hat{Y} = \sum_{i=q}^{m} U_i + \hat{V}_i$ **where** $q = 2 + m - n$ **.**

d. $\hat{Z} \approx \sum_{i=1}^{m} C_{i,m+1-i}\ \hat{f}_{m+1-i}^2 \cdots \hat{f}_{k-1}^2 (\hat{s}_k^2 / \hat{b}_k) \hat{h}_{k+1}^2 \cdots \hat{h}_{n-1}^2 \sum_{g=i+1}^{m} \hat{C}_{g,m+1-i}$

Remark. To derive formulas for the parameters f_k, s_k^2, f_k^2 an expert suggested we should examine some properties of the maximum likelihood estimators The following two properties were considered relevant.

(1) Invariance Principle: if $\hat{q}$ is the maximum likelihood estimator of parameter q and if t the is a one-one function, then $t(\hat{J})$ is the maximum likelihood estimator of $t(q)$.

(2) In general $E(t(\hat{q})) \neq tE(\hat{q})$

See Kendall, the Advanced Theory Of Statistics, vol2 (3rd ed.1973) page 44.

We considered a specific example, from DeGroot, Probability and Statistics, (2nd 3d. 1986) page 349. Let X be a normal random variable with unknown mean m and variance n, and let $\hat{m}$ and $\hat{n}$ be the maximum likelihood estimators Then

(1) The maximum likelihood estimate of $\sqrt{n}$ is $\sqrt{\hat{n}}$ and the maximum likelihood estimate of m^2 is $\hat{m}^2$

(2) $E(\hat{m}^2) \neq E^2(\hat{m})$ and in fact the maximum likelihood estimator of $E(X^2)$ is $\hat{m}^2 + \hat{n}$

The Mack estimator was suggested by (1) and the L-estimator by (2). In addition the L-estimator is unbiased in the Simple Average case.

S3.2. L-Predictor is unbiased in the Simple Average Case

For the $\alpha = 0$ simple average case the L-predictor is unbiased. We have the following theorem.

Theorem 5. Let $q = 2 + m - n$ be the first accident year which is not fully developed. In the $\alpha = 0$ simple average case we can produce an unbiased predictor of $mse(R, \hat{R}) = mse(\sum_{i=q}^{m} C_{i,n}, \sum_{i=q}^{m} \hat{C}_{i,n})$ by replacing $f_k, f_k^2, g_{i,k}^2, s_k^2$ by their unbiased estimators -- $\hat{f}_k, \hat{h}_k^2, \hat{s}_k^2, \hat{g}_{i,k}^2$.

Proof. We computed $U_i = Var_D(C_{i,n})$ and we showed that $E(\hat{V}_i) = E(V_i)$ and $E(\hat{Z}_i) = E(Z_i)$ by theorems 3 and 4. We need to show that the substitution produces an unbiased estimate for U_i, V_i and Z_i in the $\alpha = 0$ case. The substitution produces an unbiased predictor of Z_i since for $\alpha = 0$ $\hat{b}_k$ is constant and the squares $\hat{f}_k^2$ are uncorrelated. Thus

$$E(\hat{f}_{m+1-i}^2 \cdots \hat{f}_{k-1}^2 / \hat{b}_k) = E(\hat{f}_{m+1-i}^2) \cdots E(\hat{f}_{k-1}^2) / \hat{b}_k)$$

The rest is similar.

S3.3. Numerical Calculations

In most cases $\hat{s}_k^2$ is small relative to $\hat{f}_k^2$ and hence $\hat{h}_k^2 \approx \hat{g}_{i,k}^2 \approx \hat{f}_k^2$ and the L-predictors and the Mack predictors give almost the same results. We will illustrate the calculations by an example. We assume all the weights $w_{i,k} = 1$.

CUMULATIVE CLAIMS $C_{i,k}$					
acc yr	12 mo	24 mo	36 mo	48 mo	60 mo
i=1	100	200	200	200	300
i=2	100	100	200	300	300
i=3	100	200	200	250	
i=4	100	100	200		
i=5	100	150			
i=6	100				

LINK RATIOS $F_{i,k} = C_{i,k+1} / C_{i,k}$				
	24/12 mo	36/24 mo	48/36 mo	60/48 mo
	k=1	k=2	k=3	k=4
i=1	2	1	1	1.5
i=2	1	2	1.5	1
i=3	2	1	1.25	
i=4	1	2		
i=5	1.5			
ave link ratios				
f (alpha=0)	1.500	1.500	1.250	1.250
f (alpha=1)	1.500	1.333	1.250	1.200

f (alpha=2)	1.500	1.200	1.250	1.154

Projected Claims $\hat{C}_{i,k}$ using Simple Average Link Ratios (alpha=0)					
C[i,j]	j=1	j=2	j=3	j=4	j=5
acc yr	12 mo	24 mo	36 mo	48 mo	60 mo
i=1	100	200	200	200	300
i=2	100	100	200	300	300
i=3	100	200	200	250	**312.50**
i=4	100	100	200	**250.00**	**312.50**
i=5	100	150	**225.00**	**281.25**	**351.56**
i=6	100	**150.00**	**225.00**	**281.25**	**351.56**
simple ave. link ratios	1.5	1.5	1.25	1.25	
sum proj. values	0.00	150.00	450.00	812.50	1328.13

Note 281.25 = 225.00*1.25

Projected Claims $\hat{C}_{i,k}$ using Weighted Average Link Ratios (alpha=1)					
C[i,j]	j=1	j=2	j=3	j=4	j=5
	12 mo	24 mo	36 mo	48 mo	60 mo
i=1	100	200	200	200	300
i=2	100	100	200	300	300
i=3	100	200	200	250	**300.00**
i=4	100	100	200	**250.00**	**300.00**
i=5	100	150	**200.00**	**250.00**	**300.00**
i=6	100	**150.00**	**200.00**	**250.00**	**300.00**
wtd ave link ratios	1.500	1.333	1.250	1.200	
sum proj. values	0.00	150.00	400.00	750.00	1200.00

Note 150.00 = 100*1.500

Projected Claims $\hat{C}_{i,k}$ using Least Squares Link Ratios (alpha=2)					
C[i,j]	J=1	j=2	j=3	j=4	j=5
	12 mo	24 mo	36 mo	48 mo	60 mo
i=1	100	180	270	350	385
i=2	110	200	300	385	425
i=3	100	200	200	250	**288.46**
i=4	100	100	200	**250.00**	**288.46**
i=5	100	150	**180.00**	**225.00**	**259.62**
i=6	100	**150.00**	**180.00**	**225.00**	**259.62**
Least square link ratios	1.500	1.200	1.250	1.154	
sum proj. values	0	150.00	360.00	700.00	1096.15

Note 259.62=225*1.154

The following chart shows the projected Ultimate under the three methods, together with Losses To Date and the Loss Reserve, which is the excess of the projected ultimate over the Losses To Date.

	Ultimate Claims	Ultimate Claims	Ultimate Claims	Losses To Date	Reserve	Reserve	Reserve
Acc yr.	alpha=0	alpha=1	alpha=2	LTD	alpha=0	alpha=1	alpha=2
i=3	312.50	300.00	288.46	250	62.50	50.00	38.46
i=4	312.50	300.00	288.46	200	112.50	100.00	88.46
i=5	351.56	300.00	259.62	150	201.56	150.00	109.62
i=6	351.56	300.00	259.62	100	251.56	200.00	159.62
Total	1328.13	1200.00	1096.15	700.00	628.13	500.00	396.15

Next we show some of the constants used to compute the standard error. Note that the $\hat{s}_k^2$ and $\hat{b}_k = \sum_{g=1}^{m-k} w_{g,k} C_{g,k}^{a}$ are denominated in "\$" for the alpha=1 case and denominated in "\$-squared" for the alpha=2 case.

		k=1	k=2	k=3	k=4
Alpha=0	Fhat	1.5	1.5	1.25	1.25
	Beta	5	4	3	2
	sigma-sq	0.250	0.370	0.063	0.130
	sigma-sq/beta	0.050	0.093	0.021	0.065
Alpha=1	Fhat	1.500	1.333	1.250	1.200
	beta ($)	500	600	600	500
	sigma-sq ($)	25.0	44.4	12.5	30.0
	sigma-sq/beta	0.050	0.074	0.021	0.060
Alpha=2	Fhat	1.5	1.2	1.25	1.154
	beta ($$)	50,000	100,000	120,000	130,000
	sigma-sq ($$)	2,500	5,333	2,500	6,923
	sigma-sq/beta	0.050	0.053	0.021	0.053

The following chart shows $\hat{g}^2_{i,k}$; $\hat{h}^2_k$; $\hat{f}^2_k$. In the Mack predictors the value $\hat{f}^2_k$ is used for $\hat{g}^2_{i,k}$; $\hat{h}^2_k$; $\hat{f}^2_k$; and f^2_k.

		k=1	k=2	k=3	k=4
Alpha=0	$\hat{f}^2_k$	2.250	2.250	1.563	1.563
	$\hat{h}^2_k$	2.200	2.157	1.542	1.498
	$\hat{g}^2_{i,k}$	2.450	2.528	1.604	1.628
Alpha=1	$\hat{f}^2_k$	2.250	1.778	1.563	1.440
	$\hat{h}^2_k$	2.200	1.704	1.542	1.380
Alpha=2	$\hat{f}^2_k$	2.250	1.440	1.563	1.331
	$\hat{h}^2_k$	2.200	1.387	1.542	1.278

The following chart shows the computation of the standard error plus some intermediate values, for both the L-predictors and the Mack-predictors for $a = 0,1,2$.

Mean Square Error - L-Predictor, Alpha=0 (Simple Average Link Ratios)						
acc yr	U[i]	Vhat [i]	Mse(C,Chat)	Zhat	Total mse	Sqrt
i=3	8,125.00	4,062.50	12,187.50	26,406.25	38,593.75	196.45
i=4	12,085.42	5,310.42	17,395.83	23,896.88	41,292.71	203.21
i=5	36,422.67	11,530.67	47,953.34	23,061.35	71,014.69	266.49
i=6	52,111.96	14,021.02	66,132.98	0.00	66,132.98	257.16
Overall	108,745.05	34,924.61	143,669.66	73,364.47	**217,034.13**	**465.87**

Mean Square Error - Mack Predictor, Alpha=0 (Simple Average Link Ratios)						
acc yr	U[i]	Vhat [i]	mse(C,Chat)	Zhat	Total mse	Sqrt
i=3	8,125.00	4,062.50	12,187.50	26,406.25	38,593.75	196.45
i=4	12,031.25	5,364.58	17,395.83	24,140.63	41,536.46	203.80
i=5	35,572.10	11,875.81	47,447.92	23,751.63	71,199.54	266.83
i=6	49,305.01	14,622.40	63,927.41	0.00	63,927.41	252.84
Overall	105,033.37	35,925.29	140,958.66	74,298.50	**215,257.16**	**463.96**

Mean Square Error - L-Predictor, Alpha=1 (Weighted Average Link Ratios)						
acc yr	U[i]	Vhat [i]	mse(C,Chat)	Zhat	Total mse	Sqrt
i=3	7,500.00	3,750.00	11,250.00	22,500.00	33,750.00	183.71
i=4	10,950.00	4,900.00	15,850.00	19,600.00	35,450.00	188.28
i=5	25,133.33	8,445.83	33,579.17	16,891.67	50,470.83	224.66
i=6	34,194.91	10,258.15	44,453.06	0.00	44,453.06	210.84
Overall	77,778.24	27,353.98	105,132.22	58,991.67	**164,123.89**	**405.12**

Mean Square Error - Mack Predictor, Alpha=1 (Weighted Average Link Ratios)						
acc yr	U[i]	Vhat [i]	mse(C,Chat)	Zhat	Total mse	Sqrt
i=3	7,500.00	3,750.00	11,250.00	22,500.00	33,750.00	183.71
i=4	11,100.00	4,950.00	16,050.00	19,800.00	35,850.00	189.34
i=5	26,100.00	8,700.00	34,800.00	17,400.00	52,200.00	228.47
i=6	36,100.00	10,700.00	46,800.00	0.00	46,800.00	216.33
Overall	80,800.00	28,100.00	108,900.00	59,700.00	**168,600.00**	**410.61**

Mean Square Error - L-Predictor, Alpha=2 (Lease Squares Link Factors)						
acc yr	U[i]	Vhat [i]	mse(C,Chat)	Zhat	Total mse	Sqrt
i=3	6,923.08	3,328.40	10,251.48	18,639.05	28,890.53	169.97
i=4	10,118.34	4,393.49	14,511.83	15,816.57	30,328.40	174.15
i=5	20,627.22	5,923.22	26,550.44	11,846.45	38,396.89	195.95
i=6	27,457.99	7,289.38	34,747.37	0.00	34,747.37	186.41
Overall	65,126.63	20,934.50	86,061.12	46,302.07	**132,363.20**	**363.82**

Mean Square Error - Mack Predictor, Alpha=2 (Least Squares link factors)						
acc yr	U[i]	Vhat [i]	Mse(C,Chat)	Zhat	Total mse	Sqrt
i=3	6,923.08	3,328.40	10,251.48	18,639.05	28,890.53	169.97
i=4	10,251.48	4,437.87	14,689.35	15,976.33	30,665.68	175.12
i=5	21,346.15	6,090.98	27,437.13	12,181.95	39,619.08	199.05
i=6	28,835.06	7,588.76	36,423.82	0.00	36,423.82	190.85
Overall	67,355.77	21,446.01	88,801.78	46,797.34	**135,599.11**	**368.24**

APPENDIX A. Repair of Greg Taylor's Proof.

Greg Taylor was troubled by the potential bias in the Mack predictor and investigated the relationship between link factors $\hat{f}_j$ and $\hat{f}_k$ for $j<k$. See Loss Reserving, (Kluwer, 2000) pages 210-218. Taylor indicated that the following Proposition proved the link factors were independent:

Proposition. (Greg Taylor's result; our proof). Under the Chain Ladder CL1-CL3 and for $\alpha \in \{0,1,2\}$ $E(\hat{f}_k \mid \hat{f}_j) = E(\hat{f}_k)$ for $j<k$.

Proof. Recall $L_k = \{C_{i,r} : 1 \le i \le m-k; 1 \le r \le k\}$. This proof uses an independence property of conditional expectation to replace $L_k \cup \{\hat{f}_j\}$ by L_k. The variable $\hat{f}_j$ is computed using $\{C_{i,r} : 1 \le i \le m-j;\ r = j, j+1\}$ but the $\{C_{i,r} : m-k \le i < m-j;\ r = j, j+1\}$ are not needed to compute $\hat{f}_k$. Thus

$$E_{\hat{f}_j}(\hat{f}_k) = E_{\hat{f}_j} E_{L_k \cup \{\hat{f}_j\}}(\hat{f}_k) \quad \text{(Property conditional expectation)}$$

$$= E_{\hat{f}_j} E_{L_k}(\hat{f}_k) \quad \text{(Using independence of accident rows)}$$

But $E_{L_k}(\hat{f}_k) = f_k$ by a prior theorem, and f_k is a constant. Thus

$$E_{\hat{f}_j}(\hat{f}_k) = E_{\hat{f}_j} E_{L_k}(\hat{f}_k) = f_k = E(\hat{f}_k)$$

Remark. The above proposition, however, does not prove $\hat{f}_k$ and $\hat{f}_j$ are independent. Stoyanov, Counterexamples in Probability, page 54, gives an example where $E(X \mid Y) = E(X)$ but X and Y are not independent. Thus from the above Proposition we cannot conclude $\{\hat{f}_k\}$ and $\{\hat{f}_j\}$ are independent. In fact we have previously shown that for fixed accident year i that $F_{i,k}$ cannot be independent of $\{C_{i,1}, F_{i,1}, \ldots, F_{i,k-1}\}$ when $\alpha = 1$ or 2. Using the same technique we prove the following.

Proposition. Assume the chain ladder hypothesis CL1-CL3 and $\alpha = 1$ or 2. Then $\hat{f}_k$ and $L_k = \{C_{i,j} : 1 \le i \le m-k;\ 1 \le j \le k\}$ cannot be independent.

Proof . If L_k and $\hat{f}_k$ are independent, then L_k and $\hat{f}_k^2$ are independent and then $E(\hat{f}_k^2 \mid L_k)$ is a constant. But $E(\hat{f}_k^2 \mid L_k) = Var(\hat{f}_k \mid L_k) + E^2(\hat{f} \mid L_k) = \mathbf{s}_k^2 / \hat{\mathbf{b}}_k + f_k^2$ (by prior theorems).

If $\mathbf{a} = 1$ or 2 then $\hat{\mathbf{b}}_k = \sum_{i=1}^{m-k} w_{i,k} C_{i,k}^{\mathbf{a}}$ is not constant.

SECTION 4- PETERSON'S METHOD.

In his text *Loss Reserving* (Ernst & Whinney, 1981) Timothy Peterson discussed a simple method of investigating variability of the loss reserves -- that is computing the reserves using the highest and the lowest link factors -- at each age of development. This method is obviously flawed -- for example tables with many rows and columns will produce very large variations as compared to tables with only a few rows and columns. As Peterson noted on page 186-187:

In many situations, two projections comparing reserves using high and low [link] factors could be extremely misleading, and cause unwarranted concern as to the level of existing uncertainty, as will be illustrated shortly.

Nevertheless, projections using high and low factors can be useful indicators of the variability in the historical loss data.

In this note we will compute the variability of loss data using a variation of Peterson's technique. Instead of using the highest and lowest link factors we use the "average" link factors plus and minus one-two standard deviations, where the standard deviation is computed using the assumptions of the chain ladder hypothesis.

The variability of the projected ultimate claims using the link factors plus (or minus) one standard deviation was about 70% of the total computed by the Mack or the L-predictors of the mean square error. The minimum-maximum link ratio method is much easier to program than the Mack or L-predictors. The data below is from Table 7-4 of the Peterson text. (We show rounded values but the actual computations used the exact data.)

Cumulative Paid Claims ($000) (From Table 7.4 in Peterson, Loss Reserving)							
	k=1	K=2	K=3	K=4	k=5	k=6	k=7
Acc yr.	12 mo	24 mo	36 mo	48 mo	60 mo	72 mo	84 mo
I=1	1,491	5,015	7,198	8,678	9,578	10,094	10,181
I=2	1,902	6,210	8,912	10,624	11,720	12,410	12,597
I=3	2,053	7,090	10,248	12,296	13,538	14,414	0
I=4	2,338	8,216	11,975	14,442	15,833	0	0
I=5	2,861	9,730	14,182	17,361	0	0	0
I=6	3,123	10,851	15,404	0	0	0	0
I=7	3,756	11,959	0	0	0	0	0
I=8	4,181	0	0	0	0	0	0

LINK RATIOS - F[i; k]						
	k=1	k=2	K=3	k=4	k=5	k=6
	24/12 mo	36/24 mo	48/36 mo	60/48 mo	72/60	84/72
I=1	3.3631	1.4353	1.2056	**1.1037**	1.0538	1.0086
I=2	3.2656	1.4352	1.1921	1.1031	1.0589	**1.0151**
I=3	3.4542	1.4454	1.1998	1.1010	**1.0647**	
I=4	**3.5148**	**1.4575**	1.2060	1.0963		
I=5	3.4006	1.4575	**1.2242**			
I=6	3.4742	1.4195				
I=7	3.1835					
f-simple ave	3.3794	1.4418	1.2055	1.1011	1.0591	1.0119
f-wtd ave	3.3708	1.4416	1.2073	1.1006	1.0598	1.0122
f-least squares	3.3571	1.4410	1.2093	1.1000	1.0603	1.0125

*Note 3.1835 = 11959 / 3756

Various constants, using alpha = 1 -- weighted average link ratios						
	k=1	K=2	k=3	k=4	k=5	k=6
fhat, wtd ave	3.3708	1.4416	1.2073	1.1006	1.0598	1.0122
fhat, min	3.1835	1.4195	1.1921	1.0963	1.0538	1.0086
fhat, max	3.5148	1.4575	1.2242	1.1037	1.0647	1.0151
fhat, variance	3.9383E-03	5.7300E-05	3.1952E-05	2.1905E-06	6.6760E-06	6.2981E-06
fhat, std dev	.0627	.007569	.056526	.001480	.002583	.002509
fhat wtd ave - std dev	3.3081	1.4341	1.2016	1.0991	1.0572	1.0097
fhat wtd ave + std dev	3.4336	1.4492	1.2130	1.1020	1.0623	1.0147

Recall that the conditional variance is: $Var(\hat{f}_k \mid L_k) = \hat{s}_k^2 / \hat{b}_k$ where $\hat{b}_k = \sum_{i=1}^{m-k} w_{i,k} C_{i,k}$ and where $L_k = \{C_{i,j} : 1 \le i \le m-k;\ 1 \le j \le k\}$. The standard deviation (std dev) is the square root of the conditional variance.

Projected Using wtd average							
	K=1	K=2	k=3	k=4	k=5	k=6	k=7
	12 mo	24 mo	36 mo	48 mo	60 mo	72 mo	84 mo
I=1	1,491	5,015	7,198	8,678	9,578	10,094	10,181
I=2	1,902	6,210	8,912	10,624	11,720	12,410	12,597
I=3	2,053	7,090	10,248	12,296	13,538	14,414	**14,589.9**
I=4	2,338	8,216	11,975	14,442	15,833	**16,779.1**	**16,983.7**
I=5	2,861	9,730	14,182	17,361	**19,107.0**	**20,248.7**	**20,495.6**
I=6	3,123	10,851	15,404	**18,596.9**	**20,466.9**	**21,689.8**	**21,954.2**
I=7	3,756	11,959	**17,240.1**	**20,813.9**	**22,906.8**	**24,275.5**	**24,571.5**
I=8	4,181	**14,095.0**	**20,320.1**	**24,532.4**	**26,999.2**	**28,612.4**	**28,961.3**
fhat, wtd ave	3.3708	1.4416	1.2073	1.1006	1.0598	1.0122	
Sum proj values		14,095	37,560	63,943	89,480	111,606	127,556

Projected - using weighted average link factors less one standard deviation.							
	k=1	k=2	k=3	k=4	k=5	k=6	k=7
i=3	2,053	7,090	10,248	12,296	13,538	14,414	**14,553.7**
i=4	2,338	8,216	11,975	14,442	15,833	**16,738.2**	**16,900.3**
i=5	2,861	9,730	14,182	17,361	**19,081.3**	**20,172.2**	**20,367.5**
i=6	3,123	10,851	15,404	**18,509.8**	**20,343.6**	**21,506.7**	**21,714.9**
i=7	3,756	11,959	**17,149.6**	**20,607.7**	**22,649.3**	**23,944.2**	**24,176.0**
i=8	4,181	**13,832.6**	**19,837.1**	**23,837.1**	**26,198.7**	**27,696.5**	**27,964.6**
fhat less one std. deviation	3.3081	1.4341	1.2016	1.0991	1.0572	1.0097	
Sum proj values		13,833	36,987	62,955	88,273	110,058	125,677

*Note 13832.6 = 4181 * 3.3081

Projected - using weighted average link factors plus one standard deviation							
	k=1	k=2	k=3	k=4	k=5	k=6	k=7
i=3	2,053	7,090	10,248	12,296	13,538	14,414	**14,626.1**
i=4	2,338	8,216	11,975	14,442	15,833	**16,820.1**	**17,067.3**
i=5	2,861	9,730	14,182	17,361	**19,132.7**	**20,325.4**	**20,624.2**
i=6	3,123	10,851	15,404	**18,684.0**	**20,590.3**	**21,873.9**	**22,195.4**
i=7	3,756	11,959	**17,330.6**	**21,021.2**	**23,166.0**	**24,610.1**	**24,971.9**
i=8	4,181	**14,357.4**	**20,807.1**	**25,238.0**	**27,813.0**	**29,546.8**	**29,981.2**
fhat plus one std. deviation	3.4336	1.4492	1.2130	1.1020	1.0623	1.0147	
Sum proj values		14,357	38,138	64,943	90,702	113,176	129,466

Note 14357.4 = 4181* 3.4336

We calculated the mean square error using the Mack predictor (and in this case the L-predictor gave almost the same value.)

Mean Square Error, Mack Predictors, Alpha=1						
acc yr	U[i]	Vhat[i]	mse(C,Chat)	Zhat[i]	Total mse	Sqrt
i=3	3,351	1,309	4,660	20,264	24,924	157.872
i=4	9,389	3,488	12,876	39,422	52,299	228.690
i=5	14,104	5,839	19,943	43,012	62,955	250.908
i=6	51,129	17,265	68,395	84,200	152,594	390.633
i=7	107,697	38,273	145,971	90,222	236,192	485.996
i=8	866,973	343,886	1,210,859	0	1,210,859	1,100.391
Overall	1,052,644	410,060	1,462,704	277,119	1,739,823	**2,614.490**

Note 2614.490 is the square root of 1,739,823.

Mean Square Error as a percent of Ultimate and of Reserves, Alpha=1						
	Ultimate	Losses To Date				
acc yr	C[i,n]	C[i,p]	Reserve[i]	Sqrt	Sqrt / Ult	Sqrt / Res
i=3	14,590	14,414	176	158	1.08%	89.84%
i=4	16,984	15,833	1,151	229	1.35%	19.88%
i=5	20,496	17,361	3,134	251	1.22%	8.01%
i=6	21,954	15,404	6,550	391	1.78%	5.96%
i=7	24,571	11,959	12,613	486	1.98%	3.85%
i=8	28,961	4,181	24,780	1,100	3.80%	4.44%
Overall	127,556	79,152	48,404	**2,614**	**2.05%**	**5.40%**

Note 48,404 = 127,556 - 79,152

Various Constants -- Alpha=1 Weighted Average						
	k=1	k=2	k=3	k=4	k=5	k=6
fhat, wtd ave	3.3708	1.4416	1.2073	1.1006	1.0598	1.0122
fhat squared	11.3625	2.0783	1.4576	1.2112	1.1231	1.0245
beta ($)	10,644.2	36,260.7	52,515.0	63,401.3	50,669.3	36,917.8
sigma-sq ($)	41.920	2.078	1.678	0.139	0.338	0.233
sigma-sq / beta	3.9383E-03	5.7300E-05	3.1952E-05	2.1905E-06	6.6760E-06	6.2981E-06
std dev (fhat)	0.0628	0.0076	0.0057	0.0015	0.0026	0.0025
hhat-sq	11.3586	2.0783	1.4575	1.2112	1.1231	1.0245

In the chart below we compare the various projected ultimate with the ultimate computed using the alpha=1 weighted average link ratios. Timothy Peterson computed the ultimate using maximum and minimum link ratios in his tables 7-11 and 7-13. The results are shown below.

Variations in Projected Ultimate					
	Projected Ultimate	Loss To Date	Reserve	Difference versus row (1)	Difference as % Ultimate
1. Average link ratio	127,556	79152	48,404	0	0.00%
2. Average link ratio plus one std dev.	129,466	79152	46,525	1,910	**1.50%**
3. Average link ratio less one std	125,677	79152	50,314	-1,879	**-1.47%**

dev					
4. Maximum link ratios (tab 7-11)	131,681	79152	52,529	4,125	3.19%
5. Minimum Link ratios (tab 7-13)	122,819	79152	43,667	-4,737	-3.77%

Mean Square Error as a percent of Ultimate and Reserves		
	Total	mse as percent
Projected Ultimate	127566	**2.05%**
Losses to Date	79,152	3.30%
Projected Reserve	48,404	5.40%
Mean Square Error	2,614	

SECTION 5. MURPHY'S VARIABILITY MEASURE.

In this section we examine the variability estimates given in Daniel Murphy's paper, Unbiased Loss Development Factors, Casualty Actuarial Forum (1994), pp. 154-222.

The Murphy Prediction Error

We follow Murphy's notation, with some modifications -- so that the notation is similar to Mack's. We let

$m+1=I+1$ = number of accident years

$n+1=J+1=$ number of development ages

$C_{i,j}$ = cumulative claims ($) for accident year $i=0,\ldots,I$ and $j=0,\ldots,J$

$C_{0,0}$ refers to the youngest accident year, and development age 12 months.

[In Mack's papers $C_{0,0}$ refers to the oldest accident year and development age 12 months.]

Typically, $I=J$ but sometimes $I>J$.

The losses to date are the diagonal entries: $\{C_{0,0},C_{1,1},\ldots,C_{J,J}\}$

The "known" values are those on or below the diagonal: $D=\{C_{i,j} : i \geq j\}$

The values that are "unknown" and must be projected are $\{C_{i,j} : i<j\}$

For $m=5, n=4$ the following are the known values

$$\begin{pmatrix} C_{0,0} & & & \\ C_{1,0} & C_{1,1} & & \\ C_{2,0} & C_{2,1} & C_{2,2} & \\ C_{3,0} & C_{3,1} & C_{3,2} & C_{3,3} \\ C_{4,0} & C_{4,1} & C_{4,2} & C_{4,3} \end{pmatrix}$$

Definitions

(0) $\{w_{i,k} : 0 \le i \le m,\ 0 \le k \le n\}$ are fixed constants -- we assume here that they all equal "1"

(1) $F_{i,k} = C_{i,k+1}/C_{i,k}$ for $k=0,\ldots,J-1$

(2) $\hat{f}_k = \sum_{i=k}^{m} F_{i,k} w_{i,k} / \sum_{i=k}^{m} w_{i,k}$ $(k = 0,...,n-1)$ (simple average, $a = 0$)

(3) $\hat{s}_k^2 = \frac{1}{m-k-1} \sum_{i=k}^{m} w_{i,k} (F_{i,k} - \hat{f}_k)^2$ (spread factor)

(5a) The “losses to date” are those that lie on the diagonal $LTD = \{C_{i,j} : i = j$ for i ³ $0\}$.

(5b) The ultimate losses are those on the last column $ULT = \{C_{i,n} : i = 0,...,m\}$

(5c) The unknown remaining losses are $R_i = ULT_i - LTD_i = C_{i,n} - C_{i,i}$

(6a) Hat notation for claims in the known region: $\hat{C}_{i,k} = C_{i,k}$ where i ³ k

(6b) Estimated Claims in the unknown region: $\hat{C}_{i,k} = C_{i,i} \hat{f}_i \cdots \hat{f}_{k-1}$ where $i<k$

(6c) Estimated Ultimate Claims, accident year i: $\hat{U}LT = \{\hat{C}_{i,n} : i = 0,...m\}$ where $\hat{C}_{i,n} = C_{i,i} \hat{f}_i \cdots \hat{f}_{n-1}$ where $i < n$ and $\hat{C}_{i,n} = C_{i,n}$ for i ³ n

(6d) Estimated Loss Reserve, accident year i (IBNR): $\hat{R}_i = ULT_i - LTD_i$.

(6e) Estimated Loss Reserve, all accident years $\hat{R} = \sum_{i=1}^{m} \hat{R}_i$.

We assume that the claims are “fully developed” for i ³ n. Murphy uses the notation "S_n" for unknown and not fully developed future claims and "$\hat{M}_n$" for the estimated ultimate claims.

(5b) $S_n = C_{0,n} + C_{1,n} + \cdots + C_{n-1,n} =$ unknown ultimate claims at “age n”.

(6c) $\hat{M}_n = \hat{C}_{0,n} + \hat{C}_{1,n} + \cdots + \hat{C}_{n-1,n}$ = estimated ultimate claims

Stochastic Assumptions.

We will limit our discussion to Murphy’s “simple average development” (or SAD) model, or Model III. The other models are inconsistent. [In Mack’s notation $a = 0$.] Assumptions CL1, Cl2, Cl3 are as in Mack (see Murphy page 157-158)

CL1. $E(F_{i,k} | C_{i,0},..., C_{i,k}) = f_k$ where $(i = 0,..., m)(k = 1,..., n-1)$

CL2. $Var(F_{i,k} | C_{i,0},..., C_{i,k}) = s_k^2$ where $(i = 0,..., m)(k = 1,..., n-1)$ and

CL3. The accident years $\{C_{g,0},...., C_{g,n}\}$ and $\{C_{i,0},...., C_{i,n}\}$ are independent for all g and i.

In addition we add another assumption CL4, which Murphy calls CLIA (Chain Ladder Independence Assumption; see Murphy 161)

CL4. The $\{F_{i,k} : 1 \leq k \leq n\}$ are independent.

The assumption CL4 (CLIA) is consistent CL1-CL3 with Murphy’s SAD model but inconsistent with the CL1-CL3 assumptions of his other models. Murphy uses contingent probability theory using the following sets

$D = \{C_{i,k} : i \geq k\}$ = known values
$B_k = \{C_{i,j} : j \leq k\}$ = data on or prior to development age k.
$A_i = \{C_{i,k} : k = 0,...,n\}$ = i-th accident year .
$D_i = D \cap A_i$ = i-th accident year, but known values.
$G_{i,k} = A_i \cap B_k = \{C_{i,j} : j \leq k\}$ = i-th accident year, on or prior to development age k.
The $\{B_k, A_i, G_{i,k}\}$ include both known and unknown values.

Definition. Murphy's estimate of the Prediction Error is

$$\text{PredictionError} = \text{Var}_D(S_n) + \text{Var}(\hat{M}_n).$$

Summation Theorem $\text{Var}_D\{\sum_{i=0}^{n-1} C_{i,n}\} = \sum_{i=0}^{n-1} \text{Var}_D(C_{i,n})$

Proof. See the earlier proof. Note that the independence of $\{C_{i,n} : i = 0,....,m\}$ does not imply that their projections $\{E_D(C_{i,n}) : i = 0,....,m\}$ are independent.

Our Theorem 4C (cf Murphy at 211). Assumptions as above $\text{Var}_D(C_{i,n}) = s_{n-1}^2 E_D^2(C_{i,n-1}) + (s_{n-1}^2 + f_{n-1}^2)\text{Var}_D(C_{i,n-1})$ for $i = 0,...,n-1$

Proof. Recall $D = \{C_{h,j} : j \leq h\}$ are the "known" values; $A_i = \{C_{i,0},...,C_{i,J}\}$ is the i-th accident year; $B_k = \{C_{h,j} : j \leq k\}$ are values for k-th development age and before; $D_i = D \cap A_i$; $G_{i,k} = A_i \cap B_k$. Then

$\text{Var}_D(C_{i,n}) = \text{Var}_{D_i}(C_{i,n})$ (properties of the conditional expectation and CL3)

$= E_{D_i} \text{Var}_{G_{i,n-1}}(C_{i,n}) + \text{Var}_{D_i} E_{G_{i,n-1}}(C_{i,n})$ (property of conditional expectation since $D_i \subseteq G_{i,n-1}$.)

$= E_{D_i}(C_{i,n-1}^2 s_{n-1}^2) + \text{Var}_{D_i}(f_{n-1} C_{i,n-1})$ (by CL1 and CL2)

$= s_{n-1}^2 + f_{n-1}^2 \text{Var}_{D_i}(C_{i,n-1})$ (factoring out constants)

Note $\text{Var}_{D_i}(X) = E_{D_i}^2(X) + E_{D_i}(X^2)$ for any random variable X so the result follows.

Our Theorem 7C (cf Murphy 218). Recall $S_n = \sum_{i=0}^{n-1} C_{i,n}$.

Under the Chain Ladder hypothesis CL1-CL4 we have

(a) $\text{Var}_D(S_1) = C_{0,0}^2 \ s_0^2$

(b) $\text{Var}_D(S_n) = \sum_{i=0}^{n-1} E_D^2(C_{i,n}) + (s_{n-1}^2 + f_{n-1}^2)\text{Var}_D(S_{n-1})$

Proof of (a) .

Note that $D_0 = \{C_{0,0}\}$ and $S_1 = C_{0,0}$ is the only "known" value for accident year 0. Thus

$$\mathrm{Var}_D(S_1) = \mathrm{Var}_{C_{0,0}}(C_{0,1}) = s_0^2\, C_{0,0}^2 \text{ (by CL1)}$$

Proof of (b) By the **Summation Theorem** $\mathrm{Var}_D(S_n) = \mathrm{Var}_D\{\sum_{i=0}^{n-1} C_{i,n}\} = \sum_{i=0}^{n-1} \mathrm{Var}_D(C_{i,n})$.

Hence by theorem 4C.

$$\mathrm{Var}_D(S_n) = \sum_{i=0}^{n-1} \mathrm{Var}_D(C_{i,n}) = s_{n-1}^2 \sum_{i=0}^{n-1} E_D^2(C_{i,n-1}) + (s_{n-1}^2 + f_{n-1}^2) \sum_{i=0}^{n-1} \mathrm{Var}_D(C_{i,n-1}).$$

Now $\mathrm{Var}_D(C_{n-1,n-1}) = 0$ since $C_{n-1,n-1} \in D$, so the second sum is $\mathrm{Var}_D(S_{n-1})$.

===

Lemma 6. 1 Recall $\hat{M}_n = \hat{C}_{0,n} + \hat{C}_{1,n} + \cdots + \hat{C}_{n-1,n}$. Then

(a) $\hat{M}_1 = \hat{C}_{0,1} = \hat{f}_0\, C_{0,0}$

(b) $\hat{M}_2 = \hat{f}_1(\hat{M}_1 + C_{1,1}) = \hat{f}_1(\hat{f}_0\, C_{0,0} + C_{1,1})$

(c) $\hat{M}_n = \hat{f}_{n-1}(\hat{M}_{n-1} + C_{n-1,n-1})$

Proof (a) Follows from the definition

Proof (b) $\hat{M}_2 = \hat{C}_{0,2} + \hat{C}_{1,2} = \hat{f}_1\, \hat{f}_0 C_{0,0} + \hat{f}_1 C_{1,1} = \hat{f}_1(\hat{M}_1 + C_{1,1})$.

Proof (c)

$$\hat{M}_n = \hat{C}_{0,n} + \hat{C}_{1,n} + \cdots + \hat{C}_{n-1,n} = \hat{f}_{n-1}(\hat{C}_{0,n-1} + \cdots \hat{C}_{n-2,n-1}) + \hat{f}_{n-1}\hat{C}_{n-1,n-1} =$$

$$= \hat{f}_{n-1}(\hat{M}_{n-1} + C_{n-1,n-1})$$

Lemma 6.2 Using CL4 we can show**:**

(a) $\hat{f}_0$ is independent of $C_{0,0}$

(b) $\hat{f}_1$ is independent of $(\hat{M}_1 + C_{1,1}) = (\hat{f}_0\, C_{0,0} + C_{1,1})$

(c) $\hat{f}_{n-1}$ is independent of $(\hat{M}_{n-1} + C_{n-1,n-1})$

Proof. (a) $\hat{f}_0$ depends on accident years 1 to m while $C_{0,0}$ is from accident year 0. The result follows from CL3 -- which says the accident years are independent.

Proof (b) By CL4 the link ratios $\hat{f}_1$ and $\hat{f}_0$ are independent. Also $\hat{f}_1$ depends on accident years 2 to m while $C_{0,0}$ and $C_{1,1}$ depend on accident years 0 and 1.

Proof (c) Follows as in (b)

Lemma 6.3. Suppose X and Y are independent random variables on the same probability space. Then $\mathrm{Var}(X\,Y) = E^2(X)\mathrm{Var}(Y) + E^2(Y)\mathrm{Var}(X) + \mathrm{Var}(X)\mathrm{Var}(Y)$

Easy Proof. Use $\mathrm{Var}(X) = E(X^2) - E^2(X)$ and expand both sides.

Theorem 6. cf Murphy p.215

$$\mathrm{Var}(\hat{M}_n) = \mathrm{Var}(\hat{f}_{n-1})\mathrm{Var}(\hat{M}_{n-1} + C_{n-1,n-1}) + \mathrm{Var}(\hat{f}_{n-1})E(\hat{M}_{n-1} + C_{n-1,n-1}) +$$
$$+E(\hat{f}_{n-1})\mathrm{Var}(\hat{M}_{n-1} + C_{n-1,n-1})$$

Proof. Follows immediately from Lemma 6.2 and Lemma 6.3

Murphy's Computation Rules. The $\mathrm{Var}(\hat{M}_n)$ is computed using the following "computation rules"

(a) $\mathrm{Var}(\hat{f}_{n-1})$ is estimated by $\hat{\sigma}^2_{n-1}$

(b) $E(\hat{f}_{n-1})$ is estimated by the observed value of $\hat{f}_{n-1}$

(c) $\mathrm{Var}(\hat{M}_{n-1} + C_{n-1,n-1})$ is estimate by $\mathrm{Var}(\hat{M}_{n-1})$

(d) $E(\hat{M}_{n-1} + C_{n-1,n-1})$ is estimated by the observed value of $(\hat{M}_{n-1} + C_{n-1,n-1})$

Note that theorems 6 and 7C provide the formulas for computing Murphy's "prediction error"

Var(Prediction) = $\mathrm{Var}_D(S_n) + \mathrm{Var}(\hat{M}_n)$.

Note that $\hat{f}_k,\ C_{k,k},$ and $\hat{M}_{k+1} \in$ σ-algebra generated by $D = \{C_{i,j} : i \geq j\}$ for all $k = 0,1,....,n$. Hence $E_D(\hat{M}_n) = \hat{M}_n$ and $\mathrm{Var}_D(\hat{M}_n) = 0$.

BIBLIOGRAPHY

Math books

- Billingsley, Patrick, Probability and Measure (Wiley, 3rd ed. 1995)
- Chow and Teicher, Probability Theory (Springer Verlag, 3rd ed. 1997)
- Doob, J.L., Measure Theory. Graduate Texts in Mathematics, (Springer Verlag, 1994)
- Halmos, Paul, Measure Theory (Van Nostrand, 1950)
- Loeve, Michel, Probability Theory, Graduate Texts in Mathematics (Springer Verlag, 4th ed, 1977)
- Stoyanov, Jordan, Counterexamples in Probability (Wiley, 1987)

References to standard deviation of the chain ladder.

- Mack, Thomas, (1993) Distribution-free Calculations of the Standard Error of Chain Ladder Reserve Estimates (Astin Bulletin 23 ,pp 213-225)
- Mack, Thomas, Measuring the Variability of the Chain Ladder Reserve Estimates, Casualty Forum (1994)

- Mack, Thomas, (1999) “The Standard Error of Chain Ladder Reserve Estimates” (Astin Bulletin, 29) pp 361-366.

- Murphy, Daniel M. Unbiased Los Development Factors, 1994 Casualty Actuarial Forum, pp 154-222

- Taylor, Greg (2000) Loss Reserving - An Actuarial Perspective (Kluwer), pp 211-221

About the Author:

Thomas G. Kabele
Author is a Ph.D. in Mathematics from Northwestern University in Evanston, Illinois and is a fellow of the Scoiety of Actuaries. He has 30 years of experience in the insurance industry. He was Sr. Vice President in charge of reinsurance and taxation at Guardian Life. The reinsurance profit center made over $200 millin of profits. On the taxside Mr. Kabele was responsible for lobbying, compliance and tax planning. Since 2001 Mr. Kabele has been consulitng actuary and done work in merger and aquistions and international consulting assignemnts in Kosovo, Armenia, Serbia (wice) and India (3rd time). He has written nuerous papers on taxation and reinsurance and accounting issues. He has written chapters on Life Reinsurance for Strain text on reinsurance and on reinsurance accounting (IASA text).