**Zeitschrift:** Mitteilungen / Schweizerische Aktuarvereinigung = Bulletin / Association Suisse des Actuaires = Bulletin / Swiss Association of Actuaries

**Herausgeber:** Schweizerische Aktuarvereinigung

**Band:** - (1999)

**Heft:** 2

**Rubrik:** Kurzmitteilungen

# D. Kurzmitteilungen

BJØRN SUNDT, Oslo, and OKECHUKWU EKUMA, Winnipeg

## The De Pril transform of a compound $\mathcal{R}_k$ distribution

### 1 Introduction

**1A.** The De Pril transform was defined by Sundt (1995) for probability distributions on the non-negative integers with a positive probability at zero. He discussed several properties of the De Pril transform. In particular, he derived a recursion for the De Pril transform of distributions in the classes $\mathcal{R}_k$ studied by Sundt (1992). He also derived an expression for the De Pril transform of a compound distribution expressed by convolutions of the severity distribution and the De Pril transform of the counting distribution. Unfortunately, evaluation of this expression would normally be rather time-consuming. In the present paper we shall consider the special case when the counting distribution belongs to $\mathcal{R}_k$. By applying the recursion for De Pril transforms of distributions in $\mathcal{R}_k$ and a recursion derived by Sundt (1992) for a compound distribution with counting distribution in $\mathcal{R}_k$, we shall deduce a recursion for the De Pril transform of the compound distribution. When $k$ is small, this recursion could be less time-consuming than the expression given by Sundt (1995). This also implies that in this case one would be less inclined to apply approximations instead of exact evaluation.

**1B.** In Section 2 we briefly recapitulate some results on the De Pril transform and the classes $\mathcal{R}_k$ from Sundt (1992, 1995) and present our new recursion. In Section 3 we discuss this recursion in the special case when the counting distribution belongs to $\mathcal{R}_1$. Section 4 is devoted to the question whether to apply exact or approximate evaluation of convolutions of compound distributions. This question is further discussed in Section 5 in a special case where we want to evaluate the aggregate claims distribution of a heterogeneous portfolio.

**1C.** In the present paper we shall represent a distribution by its probability function, and hence we shall normally mean its probability function when talking about a distribution.

## 2 General theory

**2A.** We shall say that a probability distribution on the non-negative integers with a positive mass at zero is $R_k[a, b]$ if there exist functions $a$ and $b$ such that its probability function $p$ satisfies the recursion

$$p(n) = \sum_{i=1}^{k} \left( a(i) + \frac{b(i)}{n} \right) p(n-i) \qquad (n = 1, 2, \dots)$$

with $p(n) = 0$ for all $n < 0$. For simplicity we shall always let $a(i) = b(i) = 0$ for all $i > k$. We shall call the class of all such distributions with a fixed $k$ $\mathcal{R}_k$. These classes were introduced by Sundt (1992), who discussed several of their properties.
We obviously have that $\mathcal{R}_{k-1} \subset \mathcal{R}_k$ for all $k$. The class $\mathcal{R}_\infty$ consists of all distributions on the non-negative integers with a positive mass at zero.

**2B.** For any distribution $f \in \mathcal{R}_\infty$, there exists a unique function $\varphi_f$ on the positive integers such that $f$ can be represented as $R_\infty[0, \varphi_f]$. This is seen by solving the recursion

$$f(x) = \frac{1}{x} \sum_{y=1}^{x} \varphi_f(y) f(x-y) \qquad (x = 1, 2, \dots) \tag{2.1}$$

for $\varphi_f(x)$ to obtain

$$\varphi_f(x) = \frac{1}{f(0)} \left( x f(x) - \sum_{y=1}^{x-1} \varphi_f(y) f(x-y) \right); \quad (x = 1, 2, \dots) \tag{2.2}$$

in this paper we shall interpret $\sum_{i=s}^{t} v_i = 0$ when $s > t$. Sundt (1995) called $\varphi_f$ the *De Pril transform* of $f$. He studied its properties and argued that it can be a useful tool for recursive evaluation of distributions in $\mathcal{R}_\infty$.
The De Pril transform is additive in the sense that if $f_1, f_2, \dots, f_m \in \mathcal{R}_\infty$, then

$$\varphi_{*_{j=1}^{m} f_j} = \sum_{j=1}^{m} \varphi_{f_j}. \tag{2.3}$$

Sundt (1995) showed that if $p$ is $R_k[a, b]$, then

$$\varphi_p(n) = n a(n) + b(n) + \sum_{i=1}^{k} a(i) \varphi_p(n-i) \qquad (n = 1, 2, \dots) \tag{2.4}$$

with $\varphi_p(n) = 0$ for all negative $n$.

**2C.** Let $\mathcal{P}_+$ denote the class of distributions on the positive integers. The compound distribution $p \vee h$ with counting distribution on the non-negative integers and severity distribution $h \in \mathcal{P}_+$ is given by

$$(p \vee h)(x) = \sum_{n=0}^{x} p(n) h^{n*}(x) . \qquad (x = 0, 1, 2, \dots)$$

We immediately see that if $p \in \mathcal{R}_\infty$, then $p \vee h \in \mathcal{R}_\infty$.
Sundt (1992) showed that if $p$ is $R_k[a, b]$, then $p \vee h$ is $R_\infty[c, d]$ with

$$\begin{aligned} c(x) &= \sum_{y=1}^{k} a(y) h^{y*}(x) ; \\ & \qquad\qquad (x = 1, 2, \dots) \\ d(x) &= x \sum_{y=1}^{k} \frac{b(y)}{y} h^{y*}(x) . \end{aligned} \tag{2.5}$$

In particular, with $k = \infty$, $a = 0$, and $b = \varphi_p$, we obtain that for every $p \in \mathcal{R}_\infty$ we have that

$$\varphi_{p \vee h}(x) = x \sum_{y=1}^{x} \frac{\varphi_p(y)}{y} h^{y*}(x) \qquad (x = 1, 2, \dots) \tag{2.6}$$

as $h^{y*}(x) = 0$ for all $y > x$. The relation (2.6) was shown by Sundt (1995).

**2D.** Let us now assume that $p$ is $R_k[a, b]$. By (2.4) with $c$ and $d$ given by (2.5) we obtain that for $x = 1, 2, \dots$

$$\varphi_{p \vee h}(x) = x c(x) + d(x) + \sum_{y=1}^{x-1} c(y) \varphi_{p \vee h}(x - y) ,$$

that is,

$$\begin{aligned} \varphi_{p \vee h}(x) = x \sum_{y=1}^{k} \left( a(y) + \frac{b(y)}{y} \right) h^{y*}(x) \\ + \sum_{y=1}^{x-1} \varphi_{p \vee h}(x - y) \sum_{z=1}^{k} a(z) h^{z*}(y) . \end{aligned} \tag{2.7}$$

By letting $k = \infty$, $a = 0$, and $b = \varphi_p$ in (2.7) we obtain (2.6).
We see that in particular for low values of $k$ (2.7) is much less time-consuming than (2.6) as in (2.7) one would need $h^{i*}$ only for $i = 1, \dots, k$.

## 3 The case $k = 1$

**3A.** Let us now consider the special case $k = 1$. Then (2.4) gives

$$\varphi_p(n) = (a+b)a^{n-1}. \qquad (n = 1, 2, \dots) \tag{3.1}$$

Furthermore, (2.7) reduces to

$$\varphi_{p\vee h}(x) = (a+b)xh(x) + a\sum_{y=1}^{x-1} h(y)\varphi_{p\vee h}(x-y)\,, \qquad (x = 1, 2, \dots) \tag{3.2}$$

whereas insertion of (3.1) in (2.6) gives

$$\varphi_{p\vee h}(x) = x(a+b)\sum_{y=1}^{x} \frac{a^{y-1}}{y} h^{y*}(x)\,. \qquad (x = 1, 2, \dots) \tag{3.3}$$

It is well known that $R_1[a, b]$ is binomial if $a < 0$. Poisson if $a = 0$, and negative binomial if $a > 0$ (cf. Sundt & Jewell (1981)). Let us consider (3.1) and (3.2) in each of these three cases.

**3B.** *Binomial with parameters* $(t, \pi)$.

$$p(n) = \binom{t}{n}\pi^n(1-\pi)^{t-n}. \quad (n = 0, 1, \dots, t;\ t = 1, 2, \dots;\ 0 < \pi < 1)$$

Then

$$a = -\frac{\pi}{1-\pi}; \qquad b = (t+1)\frac{\pi}{1-\pi}$$

$$\varphi_p(n) = -t\left(\frac{\pi}{\pi-1}\right)^n \qquad (n = 1, 2, \dots) \tag{3.4}$$

$$\varphi_{p\vee h}(x) = \frac{\pi}{1-\pi}\left(txh(x) - \sum_{y=1}^{x-1} h(y)\varphi_{p\vee h}(x-y)\right). \quad (x = 1, 2, \dots) \tag{3.5}$$

Any distribution $f \in \mathcal{R}_\infty$ can be represented in the form $p \vee h$ where $p$ is a Bernoulli distribution with parameter

$$\pi = 1 - p(0)\,,$$

(that is, a binomial distribution with parameters $(1, \pi)$) and $h \in \mathcal{P}_+$ is given by

$$h(x) = \frac{f(x)}{\pi}\,. \quad (x = 1, 2, \dots)$$

Insertion in (2.2) gives

$$\varphi_f(x) = \frac{\pi}{1-\pi}\left(xh(x) - \sum_{y=1}^{x-1} h(y)\varphi_f(x-y)\right), \qquad (x = 1, 2, \dots)$$

that is, we obtain (3.5) with $t = 1$.
In general we can write (3.5) as

$$\frac{\varphi_{p\vee h}(x)}{t} = \frac{\pi}{1-\pi}\left(xh(x) - \sum_{y=1}^{x-1} h(y)\frac{\varphi_{p\vee h}(x-y)}{t}\right), \quad (x = 1, 2, \dots)$$

which immediately follows from the Bernoulli case and (2.3) as the compound binomial distribution is the $t$-fold convolution of the corresponding compound Bernoulli distribution.
In the Bernoulli case with $t = 1$, (3.5) was given in formula (2) in De Pril (1989).

**3C.** *Poisson distribution with parameter* $\lambda$.

$$p(n) = \frac{\lambda^n}{n!}e^{-\lambda}. \qquad (n = 0, 1, \dots ; \lambda > 0)$$

Then

$$a = 0; \qquad b = \lambda \tag{3.6}$$

$$\varphi_p(n) = \begin{cases} \lambda & (n = 1) \\ 0 & (n = 2, 3, \dots) \end{cases} \tag{3.7}$$

$$\varphi_{p\vee h}(x) = \lambda x h(x); \qquad (x = 1, 2, \dots) \tag{3.8}$$

the latter formula is also obtained by inserting (3.6) in (3.3).

**3D.** *Negative binomial distribution with parameters* $(\alpha, \pi)$.

$$p(n) = \binom{\alpha + n - 1}{n}(1-\pi)^\alpha \pi^n. \quad (n = 0, 1, \dots ; \alpha > 0; 0 < \pi < 1)$$

Then

$$a = \pi; \qquad b = (\alpha - 1)\pi \tag{3.9}$$

$$\varphi_p(n) = \alpha\pi^n \qquad (n = 1, 2, \dots)$$

$$\varphi_{p\vee h}(x) = \pi\left(\alpha x h(x) + \sum_{y=1}^{x-1} h(y)\varphi_{p\vee h}(x-y)\right). \quad (x = 1, 2, \dots) \tag{3.10}$$

Insertion of

$$\lambda = -\ln p(0) = \alpha|\ln(1-\pi)|$$
$$k(x) = \frac{\varphi_{p\vee h}(x)}{\lambda x} \quad (x = 1, 2, \dots) \tag{3.11}$$

in (3.10) and division by $\lambda x$ gives

$$k(x) = \pi\left(\frac{h(x)}{|\ln(1-\pi)|} + \sum_{y=1}^{x-1}\left(1 - \frac{y}{x}\right)h(y)k(x-y)\right). \qquad (x = 1, 2, \dots)$$

This is the recursion presented by Sundt & Jewell (1981) (apart from a misprint in that paper) for the compound distribution $q \vee h$, where $q$ is the logarithmic distribution given by

$$q(x) = \frac{1}{|\ln(1-\pi)|}\frac{\pi^x}{x}. \qquad (x = 1, 2 \dots)$$

Thus $k = q \vee h$. From (3.8) and (3.11) we conclude that $p \vee h = r \vee k$, where $r$ is the Poisson distribution with parameter $\lambda$. Thus

$$p \vee h = r \vee (q \vee h) = (r \vee q) \vee h\,.$$

In particular, when $h$ is concentrated in one, we obtain that $p = r \vee q$. This representation of a negative binomial distribution as a compound Poisson distribution with a logarithmic severity distribution was presented independently by Ammeter (1949) and Quenouille (1949).

**3E.** To summarise, in the binomial case the recursion (3.2) gives simply a reformulation of the general recursion (2.2) for the De Pril transform, in the Poisson case it is trivial, and in the negative binomial case the recursion has earlier been deduced within another context. Thus, with counting distributions in $\mathcal{R}_1$ (2.7) does not bring much new. However, it gives a unification.

## 4 Exact evaluation or approximation?

**4A.** To cover approximations to distributions when the approximations are not necessarily distributions themselves, Dhaene & Sundt (1998) extended the definition of the De Pril transform to the class $\mathcal{F}_0$ of functions on the non-negative integers with a positive mass at zero. Formula (2.7) is easily generalised to that situation.

**4B.** By combining (2.3) and (2.6) we obtain that if $p_j \in \mathcal{R}_\infty$ and $h_j \in \mathcal{P}_+$, then

$$\varphi_{*_{j=1}^m (p_j \vee h_j)}(x) = \sum_{j=1}^m \varphi_{p_j \vee h_j}(x)$$

$$= x \sum_{y=1}^x \frac{1}{y} \sum_{j=1}^m \varphi_{p_j}(y) h_j^{y*}(x)\,. \quad (x = 1, 2, \dots) \tag{4.1}$$

When $x$ gets large, evaluation of this formula becomes rather time-consuming. Therefore Dhaene & Sundt (1998) suggested as an approximation to replace $\varphi_{p_j}(y)$ by zero when $y$ is greater than some integer $r$. They also discussed error bounds for such approximations. Such approximations are in particular interesting when $\varphi_{p_j}(y)$ rapidly approaches zero when $y$ increases. Such approximations were studied by De Pril (1989) in the special case when the $p_j$'s are Bernoulli distributions (cf. subsection 3B), and we shall therefore call them *De Pril approximations*.

When deciding whether to apply an approximation or an exact method, computation time should be considered against the need for accuracy. Analogous considerations would be needed when deciding the order $r$ of the approximation.

As pointed out at the end of subsection 2D, (2.7) could be much less time-consuming than (2.6) when $p_j \in \mathcal{R}_k$ with a small $k$. Thus, in that case (2.7) would make exact evaluation more attractive.

**4C.** Dhaene & Sundt (1998) in particular discussed the case when $p_j$ is $R_1[a_j, b_j]$. Obviously the approximation is interesting only when $a_j \neq 0$, that is, when $p_j$ is either binomial or negative binomial. However, we also do not want $a_j$ to be too far from zero as we want $\varphi_{p_j}(y)$ to rapidly approach zero when $y$ increases.

Sundt & Jewell (1981) showed that we always have $a_j < 1$. For $a_j \leq -1$, $\varphi_{p_j}(y)$ diverges when $y \uparrow \infty$. From the discussion on $\mathcal{R}_1$ in Section 3 follows that it occurs only in the binomial case with $\pi_j \geq 1/2$.

Not surprisingly, the error bounds discussed by Dhaene & Sundt (1998) increase when $|a_j|$ increases. When $|a_j| \uparrow 1$, the error bounds go to infinity.

The Bernoulli case has been discussed by De Pril (1989) in a situation where we consider $m$ independent policies over a specified period. For $j = 1, \dots, m$, we let $\pi_j$ denote the probability that the aggregate claim amount of policy $j$ is positive and $h_j$ the conditional distribution of the aggregate claim amount of the policy given that the aggregate claim amount is positive. Then the unconditional aggregate claims distribution of the policy is the compound distribution $p_j \vee h_j$, where $p_j$ denotes the Bernoulli distribution with parameter $\pi_j$. The De Pril

approximation seems reasonable when the probabilities of non-zero claims are small.

In the following section we shall discuss an application of the negative binomial case.

## 5 Modelling heterogeneous portfolios

**5A.** We want to evaluate the aggregate claims distribution $f$ of an insurance portfolio of $m$ independent policies over a specified period. For the moment we assume that the aggregate claim amount for policy $j$ ($j = 1, \dots, m$) has the compound distribution $p_j \vee h_j$, where $h_j \in \mathcal{P}_+$ and $p_j$ is Poisson with Poisson parameter $\lambda_j$. From Theorem 11.1 in Sundt (1999) it follows that

$$f = \underset{j=1}{\overset{m}{*}} (p_j \vee h_j) = p \vee h\,,$$

where $p$ is Poisson with parameter $\lambda = \sum_{j=1}^{m} \lambda_j$ and

$$h = \frac{1}{\lambda} \sum_{j=1}^{m} \lambda_j h_j\,.$$

Thus we can evaluate $f$ by the Panjer (1980) recursion

$$f(x) = \frac{\lambda}{x} \sum_{y=1}^{x} y h(y) f(x-y)\,, \qquad (x = 1, 2, \dots)$$

which easily follows by insertion of (3.8) in (2.1).

**5B.** Often one would expect that there are individual properties of an insurance policy that affect the risk, but are not reflected by the objective rating criteria applied. We shall assume that these properties affect only the claim numbers, not the severities. We assume that to each policy $j$ there is related a positive random variable $\Theta_j$, and that $\Theta_1, \dots, \Theta_m$ are independent and identically distributed. It is assumed that the conditional distribution of the number of claims from policy $j$ given that $\Theta_j = \theta$, is Poisson with parameter $\theta\lambda_j$. Thus the unconditional distribution $p_j$ is given by

$$p_j(n) = \frac{\lambda_j^n}{n!} \mathrm{E}\Theta_j^n e^{-\Theta_j \lambda_j}\,. \qquad (n = 0, 1, \dots) \tag{5.1}$$

Let us assume that the $\Theta_j$'s are gamma distributed with density

$$u(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}. \qquad (\theta \geq 0;\ \alpha, \beta \geq 0) \tag{5.2}$$

Then we easily get

$$\mathrm{E}\Theta_j^n e^{-\Theta_j \lambda_j} = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta+\lambda_j)^{\alpha+n}}, \qquad (n = 0, 1, \dots)$$

and by insertion in (5.1) and some manipulation we obtain

$$p_j(n) = \binom{\alpha+n-1}{n} \left(\frac{\beta}{\beta+\lambda_j}\right)^\alpha \left(\frac{\lambda_j}{\beta+\lambda_j}\right)^n, \qquad (n = 0, 1, \dots)$$

that is, $p_j$ is negative binomial with parameters $\left(\alpha, \dfrac{\lambda_j}{\beta+\lambda_j}\right)$. By (3.9) this implies that $p_j$ is $R_1[a_j, b_j]$ with

$$a_j = \frac{\lambda_j}{\beta+\lambda_j}; \qquad b_j = (\alpha-1)\frac{\lambda_j}{\beta+\lambda_j}. \tag{5.3}$$

**5C.** In the situation of subsection 5A it was fairly easy to evaluate the aggregate claims distribution $f$ of the portfolio. As the policies were independent and the aggregate claims distribution of each policy was compound Poisson, also $f$ is compound Poisson. Unfortunately it is not that simple in the negative binomial case of subsection 5B. In the restrictive case when $\lambda_j$ and $h_j$ are independent of $j$ for all $j$, $f$ would be a compound negative binomial distribution; in the general situation that would usually not be the case. One possibility would then be to for each $j$ evaluate $p_j \vee h_j$ by the Panjer (1981) recursion

$$(p_j \vee h_j)(x) = \sum_{y=1}^{x} \left(a_j + b_j \frac{y}{x}\right) h_j(y)(p_j \vee h_j)(x-y), \quad (x = 1, 2, \dots)$$

and then find $f = *_{j=1}^m (p_j \vee h_j)$ by brute force convolutions. However, it seems more efficient to evaluate the De Pril transform of $f$ either exact or approximate, and then find $f$ by the recursion (2.1).

For approximate evaluation one could apply the De Pril approximation in (4.1). As argued in subsection 4B, this approximation seems reasonable if $a_j$ is small, that is, by (5.3) when $\lambda_j$ is small. This could be interpreted as if policy $j$ has low risk exposure, that is, a small policy.

For large policies it seems more appropriate to apply exact evaluation. In that case one could for each of the policies evaluate the De Pril transform by the recursion (3.10) and then sum these De Pril transforms to obtain the De Pril

transform of $f$. Numerical evaluation applying this methodology on data from group life assurance has been carried out by Ekuma (1998). This methodology is closely related to the methodology discussed in Section 5 of Willmot & Sundt (1989).

**5D.** The parameter $\lambda_j$ can be interpreted as a measure of the risk volume of policy $j$. This interpretation becomes perhaps most clear when considering group insurances. Let us look at a simplified example. We consider a group life assurance portfolio. Policy $j$ covers the employees of firm $j$. At any time during the period that firm has $n_j$ employees (that is, we assume that if an employee dies during the period, he is immediately replaced with another). Conditional on $\Theta_j = \theta$, the lives of these employees are independent, and the conditional mortality rate of each employee is $\theta\mu$. Then the conditions of subsection 5B are fulfilled with $\lambda_j = n_j\mu$, and we see that a large value of $\lambda_j$ means a firm with many employees.

**5E.** In the present situation we have for simplicity considered the unconditional distribution of the aggregate claims distribution of the portfolio. However, it would be natural to believe that the number of claims of policy $j$ from earlier years would contain information about $\Theta_j$, so that one should rather apply the conditional distribution given the claim experience of the individual policies. As the gamma distributions constitute a conjugate class to the Poisson distribution, one would under reasonable assumptions obtain that also the conditional distribution of $\Theta_j$ given the claim experience is gamma (cf. e.g. Section 2 in Norberg (1989)). Thus the conditional claim number distribution of policy $j$ is negative binomial, and the discussion of subsection 5C is still valid.

## Acknowledgement

When this paper was written both authors were affiliated with the Department of Mathematics, University of Bergen.

A part of the present research was carried out while the first author stayed as GIO Visiting Professor at the Centre for Actuarial Studies, University of Melbourne.

## References

AMMETER, H. (1949). Die Elemente der kollektiven Risikotheorie von festen und zufallsartig schwankenden Grundwahrscheinlichkeiten. *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker* **49**, 35–95.

DE PRIL, N. (1989). The aggregate claims distribution in the individual model with arbitrary positive claims. *ASTIN Bulletin* **19**, 9–24.

DHAENE, J. and SUNDT, B. (1998). On approximating distributions by approximating their De Pril transforms. *Scandinavian Actuarial Journal*, 1–23.

EKUMA, O. (1998). *Aggregate claims distributions with emphasis on De Pril transforms.* Cand.scient. thesis. Department of Mathematics, University of Bergen.

NORBERG, R. (1989). Experience rating in group life insurance. *Scandinavian Actuarial Journal*, 194–224.

PANJER, H.H. (1980). The aggregate claims distribution and stop-loss reinsurance. *Transactions of the Society of Actuaries* **32**, 523–535.

PANJER, H.H. (1981). Recursive evaluation of a family of compound distributions. *ASTIN Bulletin* **12**, 22–26.

QUENOUILLE, M.H. (1949). A relation between the logarithmic, Poisson, and negative binomial series. *Biometrics* **5**, 162–164.

SUNDT, B. (1992). On some extensions of Panjer's class of counting distributions. *ASTIN Bulletin* **22**, 61–80.

SUNDT, B. (1995). On some properties of De Pril transforms of counting distributions. *ASTIN Bulletin* **25**, 19–31.

SUNDT, B. (1999). *An introduction to non-life insurance mathematics.* (4. ed.) Verlag Versicherungswirtschaft e.V., Karlsruhe.

SUNDT, B. and JEWELL, W.S. (1981). Further results on recursive evaluation of compound distributions. *ASTIN Bulletin* **12**, 27–39.

WILLMOT, G.E. and SUNDT, B. (1989). On evaluation of the Delaporte distribution and related distributions. *Scandinavian Actuarial Journal*, 101–113.

Bjørn Sundt
Vital Forsikring ASA
P.O. Box 250
N–1326 Lysaker
Norway

Okechukwu Ekuma
The Warren Centre for Actuarial Studies & Research
Faculty of Management
Room 268, Drake Centre
181 Freedman Crescent
Winnipeg
Manitoba R3T 5V4
Canada

ERHARD KREMER, Hamburg

# Threshold Lossreserving*

## 1 Introduction

Today loss reserving is a most important subfield of nonlife insurance mathematics. Many pages on that topic can be found in the US-book „Foundations of Casualty Actuarial Science" and several texts are dedicated to it (see e.g. Taylor (1986), Institute of Actuaries (1990)). Newer approaches to loss reserving are mostly based on stochastic concepts and models. Of great practical importance are methods derived with ideas of regression and time series analysis. Examples of papers which describe these methods are Kremer (1984), (1989), (1993a), (1993b), Renshaw (1989), Verrall (1989) and Dannenburg (1995). Though the yet existing techniques are already fairly well developed, further significant progress can be expected for the future. Some methods can still be refined as is shown with the following contribution. The ideas of threshold autoregression are introduced in the field of (mathematical) loss reserving. As important practical result a refined chain ladder technique comes out, that gives more realistic results than the classical procedure for certain development features.

## 2 The basic model

Denote with the random variable $X_{ij}$ on $(\Omega, \mathcal{A}, P)$ the total claims amount or burning cost of a (collective of) risk(s) in accident year no. $i$ with respect to its development year no. $j$. Then $X_\Delta = (X_{ij},\ j = 1, \ldots, n-i+1,\ i = 1, \ldots, n)$ is the so-called *run-off triangle* of known claims data. For the following suppose that one has the model:

$$\begin{aligned} X_{ij} &= a_{j1} + b_{j1} \cdot X_{i,j-1} + e_{ij1}/V_{ij}^{1/2}, \quad \text{if } X_{i1} \le r_j \\ &= a_{j2} + b_{j2} \cdot X_{i,j-1} + e_{ij2}/V_{ij}^{1/2}, \quad \text{if } X_{i1} > r_j \end{aligned} \tag{2.1}$$

with $i = 1, \ldots, n$, $j = 2, \ldots, n$, where $a_{jl}$, $b_{jl}$, $l = 1, 2$ are certain real parameters, $V_{ij} > 0$ is a known volume measure, $r_j$ the so-called (real) *threshold* and $e_{ijl}$, $l = 1, 2$ are (real) random variables with the assumptions:

*The paper is a shortened and improved version of a former contribution to the international ASTIN colloquium at Copenhagen (1996).

1.) $E(e_{ijl}) = 0$, $\mathrm{Var}(e_{ijl}) = \sigma_{jl}^2 \in [0, \infty)$
2.) $e_{ijl}$, $l = 1, 2$, $i = 1, 2, \ldots, n$, are independent (for given $j$)
3.) $e_{ijl}$, $X_{ik}$ are independent for all $l = 1, 2$ and $k < j$

Models of type (2.1) have been well-known by statisticians for more than ten years. They are usually called (self exiting) *threshold autoregressive models* [in short: (SE)TAR] (see e.g. Petrucelli & Woolford (1984), Tong (1983), and Chan (1993)). In loss reserving they are fairly new.

Often one is willing to assume in advance that a priori one has:

$$a_{jl} = 0\,, \quad l = 1, 2\,. \tag{2.2}$$

and furthermore

$$\sigma_{j1}^2 = \sigma_{j2}^2\,. \tag{2.3}$$

## 3 The loss reserving advice

As well-known the basic problem of loss reserving consists in predicting the unknown values $X_{ij}$, $j = n - i + 2, \ldots n$ of future development from the known run-off triangle $X_\Delta$ of past development. Using arguments similar to those in Kremer (1984) one obtains the following.

Under the model (2.1) predict $X_{ij}$, $j \geq n - i + 2$ by $\widehat{X}_{ij}$, calculated recursively according to the rule:

$$\begin{aligned} \widehat{X}_{ij} &= a_{j1} + b_{j1} \cdot \widetilde{X}_{i,j-1}\,, \quad \text{if } X_{i1} \leq r_j \\ &= a_{j2} + b_{j2} \cdot \widetilde{X}_{i,j-1}\,, \quad \text{if } X_{i1} > r_j \end{aligned}$$

where one puts:

$$\begin{aligned} \widetilde{X}_{il} &= \widehat{X}_{il}\,, \quad \text{if } l > n - i + 1 \\ &= X_{il}\,, \quad \text{if } l = n - i + 1\,. \end{aligned}$$

In practice the parameters $a_{jl}$, $b_{jl}$ and the threshold $r_j$ are unknown. One has to estimate them from the data of the run-off triangle as described in the following section.

## 4 Parameter estimation

For the sequel define:

$$\nu_{jl} = \sum_{i=1}^{n-j+1} V_{ij} \cdot 1_{M_{jl}}(X_{i1}), \quad l = 1, 2,$$

with the sets:

$$\begin{aligned} M_{jl} &= (-\infty, r_j], \quad \text{for } l = 1 \\ &= (r_j, \infty) \qquad \text{for } l = 2 \end{aligned}$$

and the indicator function:

$$\begin{aligned} 1_A(x) &= 0, \quad \text{for } x \notin A \\ &= 1, \quad \text{for } x \in A. \end{aligned}$$

Furthermore declare the following sum of squares:

$$S_{jl} = \left(\frac{1}{\nu_{jl}}\right) \cdot \sum_{i=1}^{n-j+1} V_{ij} \cdot (X_{ij} - a_{jl} - b_{jl} \cdot X_{i,j-1})^2 \cdot 1_{M_{jl}}(X_{i1})$$

for $l = 1, 2$ and with them:

$$S_j = \omega_{j1} \cdot S_{j1} + \omega_{j2} \cdot S_{j2},$$

where $\omega_{jl}$, $l = 1, 2$ are certain weights, which are allowed to depend on $r_j$ and satisfy:

$$\omega_{jl} \in [0, 1], \quad l = 1, 2.$$

The „natural“ choice of weights would be:

$$\omega_{jl} = \nu_{jl}/\nu_{j\cdot}, \quad l = 1, 2 \tag{4.1}$$

with:

$$\nu_{j\cdot} = (\nu_{j1} + \nu_{j2})$$

in case that (2.3) holds in addition. With the given notation one can define „optimal“ estimators $\widehat{r}_j$, $\widehat{a}_{jl}$, $\hat{b}_{jl}$ of $r_j$, $a_{jl}$, $b_{jl}$ $(l = 1, 2)$ simply as those values

that minimize $S_j$. The solution for that optimization problem can be given easily. It is obvious that one can restrict for $\widehat{r}_j$ on the values out of the set:

$$\mathcal{R}_j = \{X_{i1}\,,\, i = 1, \ldots, n - j + 1\}\,.$$

One can proceed as follows:
For each $r_j = r \in \mathcal{R}_j$ one calculates estimators $\widehat{a}_{jl}(r)$, $\hat{b}_{jl}(r)$ by minimizing $S_{jl}$ $(l = 1, 2)$. As *optimal estimators* $\widehat{r}_j$, $\widehat{a}_{jl}$, $\hat{b}_{jl}$ of $r_j$, $a_{jl}$, $b_{jl}$ one chooses that $r$, $\widehat{a}_{jl}(r)$, $\hat{b}_{jl}(r)$ for which $S_j$ becomes minimal.
With classical calculus one gets that given $r_j = r$ $(\in \mathcal{R}_j)$:

$$\widehat{a}_{jl}(r) = \sum_{i=1}^{n-j+1} \left(\frac{V_{ij}}{\nu_{jl}}\right) \cdot (X_{ij} - \hat{b}_{jl}(r) \cdot X_{i,j-1}) \cdot 1_{M_l}(X_{i1}$$

$$\hat{b}_{jl}(r) = \frac{\displaystyle\sum_{i=1}^{n-j+1} \left(\frac{V_{ij}}{\nu_{jl}}\right) \cdot \left(X_{ij} - \sum_{m=1}^{n-j+1} \left(\frac{V_{mj}}{\nu_{jl}}\right) \cdot X_{mj}\right) \cdot X_{i,j-1} \cdot 1_{M_l}(X_{i1})}{\displaystyle\sum_{i=1}^{n-j+1} \left(\frac{V_{ij}}{\nu_{jl}}\right) \cdot \left(X_{i,j-1} - \sum_{m=1}^{n-j+1} \left(\frac{V_{mj}}{\nu_{jl}}\right) \cdot X_{m,j-1}\right) \cdot X_{i,j-1} \cdot 1_{M_l}(X_{i1})}$$

where $M_l$ is just $M_{jl}$ with $r_j = r$.
For a priori (2.2) one gets more simple:

$$\hat{b}_{jl}(r) = \frac{\displaystyle\sum_{i=1}^{n-j+1} \left(\frac{V_{ij}}{\nu_{jl}}\right) \cdot X_{ij} \cdot X_{i,j-1} \cdot 1_{M_l}(X_{i1})}{\displaystyle\sum_{i=1}^{n-j+1} \left(\frac{V_{ij}}{\nu_{jl}}\right) \cdot X_{i,j-1}^2 \cdot 1_{M_l}(X_{i1})}$$

as estimator for $b_{jl}$ given $r_j = r$ (for $l = 1, 2$). Obviously the resulting method is handy though refined significantly.

One can prove that under a priori (2.2) and $V_{ij} = 1$ for all $i, j$ the estimators $\widehat{r}_j$, $\hat{b}_{jl}$ of $r_j$, $b_{jl}$ are consistent (for $n \to \infty$). A sketch of the proof can be found in Kremer (1996).

In case of a priori (2.2) and

$$e_{ij1} = e_{ij2}\,, \quad \text{for all } i\,,$$

one can give an adequate test statistic for testing the *hypothesis*:

$$H : b_{j1} = b_{j2}\,,$$

against the alternative:

$$K : b_{j1} \neq b_{j2} .$$

Similar to what is said in Petrucelli & Woolford (1984), p. 278 one can take a lower one-sided test with statistic:

$$T_n = \left(\widehat{S}_j(n)/s_j(n)\right)^{(n-j)/2},$$

where

$$\widehat{S}_j(n) = \sum_{l=1}^{2} \sum_{i=1}^{n-j+1} \frac{V_{ij}}{V_{\cdot j}} \left(X_{ij} - \hat{b}_{jl}(\widehat{r}_j) \cdot X_{i,j-1}\right)^2 \cdot 1_{\widehat{M}_{jl}}(X_{i1})$$

with

$$\begin{aligned} \widehat{M}_{jl} &= (-\infty, \widehat{r}_j\,], \quad \text{for } l = 1 \\ &= (\widehat{r}_j, \infty), \quad \text{for } l = 2 \end{aligned}$$

$$V_{\cdot j} = \sum_{i=1}^{n-j+1} V_{ij}$$

and

$$s_j(n) = \sum_{i=1}^{n-j+1} \left(\frac{V_{ij}}{V_{\cdot j}}\right) \cdot (X_{ij} - \hat{b}_j \cdot X_{i,j-1})^2$$

with the overall estimator

$$\hat{b}_j = \frac{\sum_{i=1}^{n-j+1} \left(\frac{V_{ij}}{V_{\cdot j}}\right) X_{ij} \cdot X_{i,j-1}}{\sum_{i=1}^{n-j+1} \left(\frac{V_{ij}}{V_{\cdot j}}\right) \cdot X_{i,j-1}^2}$$

Under certain regularity conditions $[-2 \cdot \ln(T_n)]$ is under $H$ asymptotically $\chi^2$-distributed with one degree of freedom. Consequently a senseful test for $H$ against $K$ is:

$$\varphi = 1_{\{T^* > \chi^2_{1;1-\alpha}\}}$$

with the test statistic $T^* = -2 \cdot \ln(T_n)$ and the $\alpha$-fractile $\chi^2_{1;1-\alpha}$ of the $\chi^2_1$-distribution.

In case that the test decides for $H$, one will take the $\hat{b}_j$ as estimator of $b_{j1} = b_{j2}$. This is just the classical procedure.

## 5 Numerical example

Take a priori (2.2) and $V_{ij} = 1$ for all $i, j$ and consider the run-off triangle $X_\Delta$:

| | | | | |
|---|---|---|---|---|
| 31.28 | 48.98 | 67.39 | 79.14 | 85.43 |
| 60.47 | 77.53 | 114.51 | 154.47 | |
| 33.77 | 49.39 | 62.65 | | |
| 67.06 | 95.49 | | | |
| 29.58 | | | | |

With the classical *chain-ladder* method (s. e.g. Kremer (1984)) one gets the following completion of the triangle to a rectangle:

**Completions 1:**

| | | | |
|---|---|---|---|
| | | | 166.83 |
| | | 80.46 | 86.85 |
| | 132.76 | 170.50 | 184.05 |
| 41.69 | 57.44 | 74.44 | 80.36 |

With the method of section 5 one has clearly more work than with the chain-ladder technique. One proceeds along the following steps.

| | | | |
|---|---|---|---|
| $j = 2$: | $r = 31.28$ | gives | $S_j = 12.75$ |
| | $r = 60.47$ | gives | $S_j = 18.86$ |
| | $r = 33.77$ | gives | $S_j = 11.55$ |
| | $r = 67.06$ | gives | $S_j = 20.99$ |

Obviously the third threshold is optimal.

But one has:

$$T^* = 1.79 < \chi^2_{1;0,9} = 2.71$$

what means that one decides for $b_{21} = b_{22}$ and has to take the overall estimator $\hat{b}_2 = 1.39$ for $b_2 = b_{21} = b_{22}$.

$j = 3$: $r = 31.28$ gives $S_j = 25.14$

$r = 60.47$ gives $S_j = 26.19$

$r = 33.77$ gives $S_j = 4.65$

Again the third threshold is optimal.

Now one has:

$$T^* = 3.46 > 2.71 .$$

One has to take the estimator $\hat{b}_{31} = 1.32$ for $b_{31}$ and $\hat{b}_{32} = 1.48$ for $b_{32}$ $(\neq b_{31})$.

$j = 4$: $r = 31.28$ gives $S_j = 0$

$r = 60.47$ gives $S_j = 51.42$

Trivially the first threshold is optimal.

Finally one has:

$$T^* = \infty > 2.71 .$$

As estimator of $b_{41}$ one gets $\hat{b}_{41} = 1.17$ and for $b_{42}$ $(\neq b_{41})$ one has to take the estimator $\hat{b}_{42} = 1.35$.

$j = 5$:

No optimality calculations possible. In the final step one has to apply the chain-ladder advice, meaning simply:

$$\widehat{X}_{i5} = (X_{15}/X_{14}) \cdot \widetilde{X}_{i4} = 1.08 \cdot \widetilde{X}_{i4} .$$

Choosing in each step the optimal threshold, one gets with the *threshold method* as completion of the triangle:

**Completions 2:**

| | | | |
|---|---|---|---|
| | | | 166.83 |
| | | 84.58 | 91.35 |
| | 141.33 | 190.80 | 206.06 |
| 41.10 | 54.30 | 70.59 | 76.24 |

The results of completions 1 and 2 differ considerably. The author judges the second one as being more reliable.

## 6 Final comments

The above method can be modified slightly. For each $j \geq 2$ the comparison with the threshold can be carried though with $X_{i,j-1}$ instead of $X_{i1}$. In the prediction phase one does not know $X_{i,j-1}$ for $j > n - i + 2$, one will replace it by the $\widehat{X}_{i,j-1}$ then. With this modification the method might be even more appealing. The reader is invited to apply this modification to the above numerical example. In practice one should adjust the claims data for inflation in advance.

## References

Chan, K.S. (1993) Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics*. 520–533.

Dannenburg, D. (1995) An autoregressive credibility IBNR model. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik*. 235–248.

Institute of Actuaries (1990) *Claims reserving manual*. London.

Kremer, E. (1984) A class of autoregressive models for predicting the final claims amount. *Insurance: Mathematics and Economics*. 111–119.

Kremer, E. (1989) Loss reserving by kernel regression. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*. 143–155.

Kremer, E. (1993a) Random coefficient autoregressive loss reserving. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik*. 237–240

Kremer, E. (1993b) Certain extensions of the chain-ladder technique. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*. 173–186. Correction note in the same journal in 1994.

Kremer, E. (1996) Threshold loss reserving. *Proceedings of the ASTIN colloquium at Copenhagen*.

Petruccelli, J.D. and Woolford, S.W. (1984) A threshold AR(1) model. *Journal of Applied Probability*. 270–286.

Renshaw, A.E. (1989) Chain ladder and interactive modelling (Claims reserving and GLIM). *Journal of the Institute of Actuaries*. 559–587.
Taylor, G.C. (1986) *Claims reserving in nonlife insurance*. North Holland, Amsterdam.
Tong, H. (1983) *Threshold models in non-linear time series analysis*. Lecture notes in Statistics. Springer, New York.
Verrall, R.J. (1989) Modelling claims runoff triangles with two-dimensional time series. *Scandinavian Actuarial Journal*. 129–138.

Erhard Kremer
Institut für Mathematische Stochastik
Universität Hamburg
Bundesstr. 55
D-20146 Hamburg

# Errata

**«Les premières tables suisses de mortalité»**
**de Philippe Chuard, Bulletin 1/1999**

Malheureusement de gênantes erreurs sont contenues dans le texte de l'article ci-dessus. Les renvois aux tableaux de la page 88 sont inexacts, il s'agit de renvois aux tableaux 5 et 6. A la page 97 les probabilités de décès sont désignées par des d et non par des q.

*Le Comité de Rédaction*