ACTUARIAL RESEARCH CLEARING HOUSE
1993 VOL. 3

ESTIMATION OF LONG TAILED UNPAID LOSSES FROM PAID LOSS DEVELOPMENT USING TRENDED GENERALIZED BONDY DEVELOPMENT

BY

BRADFORD S. GILE, FSA, MAAA

Submitted to

ACTUARIAL EDUCATION AND RESEARCH FUND
AERF PRACTITIONERS' AWARD
SEPTEMBER 21, 1993

ESTIMATION OF LONG TAILED UNPAID LOSSES FROM PAID LOSS DEVELOPMENT USING TRENDED GENERALIZED BONDY DEVELOPMENT

I. Generalized Bondy Development

In his paper, "Generalized Bondy Development", Mr. Alfred O. Weller describes a development pattern, for a fixed experience period, which all development ratios of length h years will satisfy

$$(1)\quad d(t+h)_h = \left[d(t)_h \right]^B \text{ ; for all } t >= c$$

where $0 < B < 1$ and

$$d(x)_h = \frac{\text{Losses Paid Through } x + h \text{ Years}}{\text{Losses Paid Through } x \text{ Years}}$$

and t is time in years since the beginning of the accident period under investigation.

Let $G(t) = \dfrac{\text{Losses Paid Through } t \text{ Years}}{\text{Ultimate Incurred Losses}}$. A continuous solution to (1) is given by (2) below:

(2) $G(t) = EXP(\ u(t))$, where EXP is the exponential function and

$$u(t) = -\frac{\ln(A)}{1 - B} \cdot B^{\left[\frac{t - c}{h}\right]}$$

It may be noted that

$$(3)\quad d(c)_h = \frac{G(c + h)}{G(c)} = EXP(u(c+h)-u(c)) = EXP(\ln(A)) = A.$$

The value G(t) is often referred to as the "completion factor" at age t. Many actuaries prefer to use the reciprocal of G(t), which is usually called the "tail factor" at age t. Clearly, the choice between the two is a matter of taste; both will produce identical results.

The completion factor concept is useful, however, in seeing that this model can only be descriptive of the TAIL of the ratio of paid to ultimate losses, because any global model would require that G(0) = 0. Now, G(t) may be thought of as a cumulative probability distribution. Therefore, its first derivative must be nonnegative. Moreover, during the accident period we would expect the second derivative of G to be positive because loss development should be increasing at an increasing rate while claims are still arising. The first derivative of G(t),

$$DG(t) = G(t) \cdot Du(t) = G(t) \cdot \frac{\ln(B)}{h} \cdot u(t) > 0,$$ and the second derivative of G(t) is

$$D^2 G(t) = G(t) \cdot u(t) \cdot \left[\frac{\ln(B)}{h} \right]^2 \cdot (1 + u(t)),$$ which can take on a POSITIVE value if, and only if, 1 + u(t) < 0.

If such a t exists, t >= c, then 1 + u(c) < 0 also, because u(t) is monotonically increasing. Thus, the second derivative of G is ALWAYS NEGATIVE unless $1 - \frac{\ln(A)}{1 - B} < 0$, or $\ln(A) > 1 - B$.

Therefore, if c is to be within the accident period, the value of $d(c)_h = A > EXP(1 - B)$ will need to be quite large unless B is close to unity. As a practical matter, then, this model will normally be used to describe development somewhat beyond the end of the accident period.

II. Trended Generalized Bondy Development

It sometimes happens that one model does not fit the loss development of all accident periods being examined. This situation may occur when varying trends or other influences cause paid loss development patterns to vary over time.

One solution is to develop different parametric values for different accident periods. Such a model does not, however, provide much insight into how variation is occurring. A better approach would seem to be to allow one or both of A and B to vary over time according to a specified model.

We have already seen that the parameter A is equal to the development ratio $d(c)_h$. If we plug t = c + h into (2) and solve for B, we get

$$(4) \qquad B = \frac{\ln(G(c+h))}{\ln(G(c))}$$

Now if we fix A, small changes in B can cause very dramatic changes in values of G(t). Suppose we have two values B1 and B2 for the parameter B with A fixed. Now we write G(t;A,B) for the function G with parameters A and B. Fix A.

If B2 > B1, then if we define r(t) by

$$r(t) = \frac{\ln(G(t;A,B2)}{\ln(G(t;A,B1)} = \frac{1 - B1}{1 - B2} \cdot \left[\frac{B2}{B1} \right]^{\frac{t - c}{h}}$$

Then $G(t;B2) = G(t;B1)^{r(t)}$, and r(t) explodes to infinity as t tends to infinity no matter how close B1 and B2 are to each other. Thus, small changes in B have a marked effect on the development and the effects vary in a complex fashion over time.

Now let us fix B and consider A2 > A1 > 1 as A values:

$$r(t) = \frac{\ln(G(t;A2,B)}{\ln(G(t;A1,B)} = \frac{\ln(A2)}{\ln(A1)} \text{ is a CONSTANT } k,$$

so that

$$G(t;A2,B) = G(t;A1,B)^{k} \text{ for all } t.$$

Thus, changes in A produce more stable results than changes in B.

For this reason, the simplest generalization to a model which varies the parameters over different accident periods will hold B constant and allow A to vary. This is the model that this paper will consider. Although models with variation in B would no doubt be intriguing, and may well yield useful results, we will now consider only models for which B is held constant.

In particular, we will assume that h = 1, accident periods are accident YEARS, and that the A parameter, denoted by A(y), varies by accident year y geometrically:

$$(5) \quad A(y) = A(1) \cdot R^{(y-1)} \; ; \text{ accident years } 1 <= y <= N$$
$$R > 1$$

Remembering that the A parameter is the initial development ratio d(c), we see that under this model the value of d(c) increases geometrically over time. If we now write $F(t,y) = G(t;A(y),B)$ with B fixed and A(y) as in (5), we have :

$$\ln(F(t,y)) = \frac{-\ln(A(y))}{1 - B} \cdot B^{t-c} = \frac{-\ln(A(1)) - (y-1) \cdot \ln(R)}{1 - B} \cdot B^{t-c}$$

and $\ln(F(t,1)) = \frac{-\ln(A(1))}{1 - B} \cdot B^{t-c}$, so we may write

$$(6) \quad \ln(F(t,y)) = \ln(F(t,1)) - \frac{y - 1}{1 - B} \cdot \ln(R) \cdot B^{t-c}$$

for all t >= c and accident years 1 <= y <= N.

Thus, given the first accident year development F(t,1) and R, (6) then determines F(t,y) for each accident year y.

In the next section, we will apply this model to some actual long-tailed loss development which does have markedly varying development ratios by accident year.

III. APPLICATION TO ACTUAL LOSS DEVELOPMENT

The paid development to be studied arises from a medical expense incurred insurance contract which provides benefits after the satisfaction of a fixed deductible for each occurrence (illness/injury). Although there is a maximum dollar benefit, there is no limit to the timing of benefit payments for a given occurrence as long as the policy remains in force.

Thus, all benefit payments are assigned an incurred date equal to the date of illness or injury. As one might expect, the loss development pattern is long; what one might NOT be prepared to expect is just HOW long the development is.

Table A below shows the historical run off of losses incurred prior to 1982:

TABLE A: HISTORICAL RUN OFF OF LOSSES INCURRED PRIOR TO 1982 THROUGH 12/31/91

CALENDAR YEAR	BENEFITS PAID ON LOSSES PRIOR TO 1982
1982	$10,323,818
1983	2,006,910
1984	1,134,964
1985	759,590
1986	696,838
1987	371,022
1988	220,861
1989	163,867
1990	93,867
1991	86,983
	$15,858,720

Unfortunately, with loss development that can easily exceed 15 years, it is very difficult to estimate ultimate losses. Traditional methods (and some unorthodox ones, too) applied to even five years' development may produce inadequate estimated ultimate losses when compared with full development. There are also significant differences in development by accident year. Table B shows the paid loss development through 12/31/91.

This table starts with the development from 12 months. It is clear that development from 12 to 24 months is very large relative to later 12 month intervals. This is a major reason why Bondy Development will fail if we start at an earlier development age than 24 months in this case.

TABLE B: PAID LOSS DEVELOPMENT AS OF YEAR END 1991

YEAR BENEFITS WERE PAID	YEAR LOSSES WERE INCURRED 1982	1983	1984	1985	1986
1982	$25,893,182	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
1983	11,061,692	30,714,940	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
1984	1,463,653	11,682,569	34,987,383	XXXXXXXXXX	XXXXXXXXXX
1985	950,355	1,671,802	12,367,435	34,177,430	XXXXXXXXXX
1986	667,582	899,834	2,262,510	14,350,459	39,549,678
1987	344,756	357,440	1,152,199	2,592,063	17,825,496
1988	411,555	640,625	1,054,086	1,347,414	3,164,217
1989	228,296	369,787	851,333	917,255	2,214,192
1990	51,414	299,928	457,786	805,917	1,372,068
1991	173,364	247,586	514,679	551,972	926,767

YEAR BENEFITS WERE PAID	YEAR LOSSES WERE INCURRED 1987	1988	1989	1990	1991
1982	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
1983	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
1984	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
1985	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
1986	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
1987	44,024,194	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
1988	20,468,624	45,843,638	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
1989	3,309,287	23,355,464	54,886,102	XXXXXXXXXX	XXXXXXXXXX
1990	2,236,870	4,784,667	26,608,917	58,786,420	XXXXXXXXXX
1991	1,468,859	2,100,111	5,254,928	34,079,112	65,350,261

Accordingly, all further discussion will be limited to development beyond 24 months, and we will use $c = 2$ years.

Table C below shows the 24 month development factors, defined as the ratio of cumulative paid losses at 36 months to cumulative paid losses at 24 months, for each of the accident years 1982 through 1989:

TABLE C: HISTORICAL 12 MONTH DEVELOPMENT FROM AGE 24 MONTHS

ACCIDENT YEAR	BENEFITS PAID AS OF: 24 MONTHS	36 MONTHS	DEVELOPMENT RATIO
1982	$ 36,954,874	$ 38,418,527	1.0396
1983	42,397,509	44,069,311	1.0394
1984	47,354,818	49,617,328	1.0478
1985	48,527,889	51,119,952	1.0534
1986	57,375,174	60,539,391	1.0551
1987	64,492,818	67,802,105	1.0513
1988	69,199,101	73,983,769	1.0691
1989	81,495,019	86,749,946	1.0645
	$447,797,203	$472,300,330	

A least squares geometric fit to these ratios which preserves the sum of the 36 month benefits is :

$$A(1982) = 1.038781$$
$$R = 1.003744$$
$$d(2,y) = A(1982) \cdot R^{y-1982}$$
$$y >= 1982$$

TABLE C clearly shows a significant upward trend in development by advancing accident year that cannot be ignored. Similar analyses at higher ages also show evidence of such a trend. Chart 1 on the next page shows how the fitted A parameters compare with the actual observed values.

It NOW remains to estimate the Bondy parameter B. This should be done using as much data as possible, because the development is VERY long tailed. Once we have a representation for F(t,y), we will be able to estimate the ULTIMATE incurred losses for each year y having at least 2 years of development :

$$\text{ULT}(y) = \frac{\text{Year } y \text{ Losses Paid Through } t \text{ Years}}{F(t,y)} \quad ; \; t >= 2$$

Thus, for each year y, we will have SEVERAL different estimates of ultimate losses. We would like to have such estimates differ as little as possible. Let DIFF(y) be the excess of the maximum over minimum ultimate loss estimate for year y, and let S be the sum of the DIFF(y) for y = 1982 to y = 1989. We then seek that value of B, $0 < B < 1$, such that S is minimized.

The value of B which minimizes S is

$$B = .642834$$

		ULTIMATE LOSS ESTIMATES		PERCENTAGE
YEAR	A	MINIMUM	MAXIMUM	DIFFERENCE
1982	1.0388	$41,108,909	$41,374,172	0.6%
1983	1.0427	46,757,339	47,659,400	1.9%
1984	1.0466	53,515,135	54,132,081	1.2%
1985	1.0505	55,376,705	55,859,079	0.9%
1986	1.0544	66,552,479	66,721,764	0.3%
1987	1.0584	74,589,677	75,595,437	1.3%
1988	1.0623	81,597,551	82,490,302	1.1%
1989	1.0663	97,377,068	97,544,613	0.2%

The small percentage variation in estimates within each accident year is, in itself, an indication that this model is reasonable.

It must be admitted, however, that the model does NOT give us ultimate losses for 1991 because as of 12/31/91, $t = 1 < c$. We get around this constraint simply by looking at the ratio of losses paid through 12 months to ultimate losses for each of the years y = 1982 to 1990:

ESTIMATION OF ULTIMATE LOSSES THROUGH 1991

YEAR	ESTIMATED ULTIMATE	PAID AS OF 12 MONTHS	OBSERVED COMPLETION	FITTED
1982	$41,374,172	$25,893,182	0.625830	0.656745
1983	47,133,983	30,714,940	0.651652	0.640832
1984	54,132,081	34,987,383	0.646334	0.625304
1985	55,577,586	34,177,430	0.614950	0.610152
1986	66,721,759	39,549,678	0.592755	0.595368
1987	74,589,677	44,024,194	0.590218	0.580942
1988	81,597,551	45,843,638	0.561826	0.566866
1989	97,377,068	54,886,102	0.563645	0.553130
1990	112,323,538	58,786,420	0.523367	0.539727
1991	124,086,792	65,350,261		**0.526650**

The fitted values are a geometric regression on the ratios of 12 month to Ultimate losses. The 12 month paid losses for 1991 divided by the fitted 1991 ratio gives us the estimated Ultimate losses for 1991.

The model thus gives us the following estimate of Unpaid Losses as of 12/31/91 on losses incurred from 1/1/82 to 12/31/91:

ULTIMATE, 1982-1991	$754,914,207
PAID TO 12/31/91	654,130,151
12/31/91 UNPAID	$100,784,057

Now, the goodness of fit of any model to past experience does not guarantee that the model is a good predictor of FUTURE experience. The only real test lies in comparing predictions with what actually subsequently happens. Fortunately, the model may be used not only to predict the unpaid losses, but it also predicts how those losses will be paid out in the future. Thus, the predicted **1992** payments on losses incurred in year y is given by:

$$\text{PAYMENT}(y,1992) = \text{ULT}(y)\cdot(\,F(1993-y,y) - F(1992-y,y)),\quad 1982 <= y <= 1991$$

These predictions are THEN compared to what actually happened in calendar year 1992:

1992 PAYMENTS BY ACCIDENT YEAR

YEAR	PREDICTED	ACTUAL	ERROR
1982	$ 45,787	$ 212,900	(167,113)
1983	88,951	212,654	(123,702)
1984	172,608	415,437	(242,829)
1985	296,810	679,320	(382,509)
1986	591,382	1,042,184	(450,802)
1987	1,085,831	1,081,057	4,774
1988	1,925,164	1,788,616	136,548
1989	3,655,475	2,770,606	884,869
1990	6,528,946	6,911,345	(382,399)
1991	36,172,947	35,512,644	660,303
	$50,563,902	$50,626,763	($62,861)

The RESULT is, in the aggregate, quite satisfying with an error of only -0.12%. This result suggests that the model is quite good for loss reserving purposes.

The reader may well note, however, that the percentage error is relatively large for losses incurred prior to 1987. In this regard, it is interesting to note that if we extend TABLE A to include 1992 payments, we see that the 1992 payments on losses incurred prior to 1982 is abnormally large:

TABLE D: HISTORICAL RUN OFF OF LOSSES INCURRED PRIOR TO 1982 THROUGH 12/31/92

CALENDAR YEAR	BENEFITS PAID ON LOSSES PRIOR TO 1982
1982	$10,323,818
1983	2,006,910
1984	1,134,964
1985	759,590
1986	696,838
1987	371,022
1988	220,861
1989	163,867
1990	93,867
1991	86,983
1992	**269,708**

Attempts have been made to "explain" the error variation by generalizing the model to variation on B. Such attempts have failed. It is probable that, as with almost all mathematical representations of the real world, forces are operating which are not readily modeled and which require separate investigation.

APPENDIX : DEVELOPMENT OF THE COMPLETION FUNCTION G(t)

We start with the basic criterion:

$$d(t+h)_h = d(t)^B \quad \text{for all } t >= c$$

Then we note that the ratio of Ultimate to Paid through time t

is

$$\frac{\text{Ultimate}}{\text{PAID}(t)} = \prod_k d(t+k\cdot h)_h = \prod_k \left(d(t)_h\right)^{\left(B^k\right)} = \left(d(t)_h\right)^{\left[\frac{1}{(1-B)}\right]}$$

where k ranges from zero to infinity.

Thus, we have

$$(1) \quad G(t)_h = \frac{\text{PAID}(t)}{\text{Ultimate}} = \left(d(t)_h\right)^{\left[-\frac{1}{(1-B)}\right]}$$

In particular, consider any integer n and note that

$$(2) \quad d(c+n\cdot h)_h = \left(d(c)_h\right)^{\left(B^n\right)}$$

We extend this functional relationship to all real z >= 0 by

$$(3) \quad d(c+z)_h = \left(d(c)_h\right)^{\left[B^{\left(\frac{z}{h}\right)}\right]}$$

This extension is NOT unique, so other formulations are possible.

Substitution of t = c +(t-c) into (1) yields:

$$(4) \quad G(t)_h = \left[d(c+(t-c))_h\right]^{\left[-\frac{1}{(1-B)}\right]} = \left(d(c)_h\right)^{\frac{-B^{\left(\frac{t-c}{h}\right)}}{(1-B)}}$$

In the paper, the formulation of G(t) is slightly different from (4) in that it expresses G(t) as a power of e :

$$(5) \quad G(t) = e^{\left[-\frac{B^{\left(\frac{t-c}{h}\right)}}{(1-B)} \cdot \ln\left(d(c)_h\right)\right]}$$

If we note that $A = d(c)_h$ **, we get the final result:**

$$(6) \quad G(t) = e^{u(t)}$$

where $u(t) = \frac{-\ln(A)}{(1-B)} \cdot B^{\left(\frac{t-c}{h}\right)}$

This formulation is unique if t = c + n·h, where n is an integer, but is NOT unique for intermediate values of t.