JIA 114 (1987) 591-599

BIAS IN q AND m RATE ESTIMATES

By L. A. Roberts, M.Sc., A.I.A.
(of the London School of Economics and Political Science).

INTRODUCTION

Bias in estimates of the initial decremental rate q and the central decremental rate m is considered for several survival functions. It is first shown that the central rate estimate $\mathring{m}$ is unbiased when the survival function is an exponential curve; $\mathring{q}$ is similarly unbiased when the survival function assumes the Balducci shape. The evidence presented here also indicates that the proportional bias (the bias divided by the true value) is approximately twice as great for $\mathring{q}$ as for $\mathring{m}$ when the survival function is a straight line.

Life tables in common use in the United Kingdom are, however, concave at ages arising most frequently in actuarial calculations, while none of the above curves is concave. In order to investigate the bias properties of general survival functions a quadratic is chosen, with a parameter indicating the extent of curvature. For this particular class of survival functions, one can show that bias is approximately minimized when exposure to risk is symmetric about the middle of the rate year.

It is shown that the ratio of proportional bias in q and m rate estimates depends primarily upon the curvature of the survival function, i.e. upon the level of convexity or concavity, and only slightly upon the distribution of exposure over the rate year. For most survival functions encountered in actuarial work the force of decrement will increase over the rate year, in which case the m estimate will suffer less bias than the q estimate. Even when the force falls slightly over the rate year, the m rate estimate may still be subject to smaller bias.

The level of bias is finally compared with the width of the confidence interval surrounding the rate estimate. The results indicate that in practice bias in estimated mortality rates will generally be negligible, except perhaps for very advanced ages and when the exposure is very uneven over the rate year. The extent of bias for decrements other than mortality is difficult to quantify because of the absence of standard tables from which to obtain an idea of the degree of concavity or convexity of the survival function at the age considered.

The plan of the paper is as follows. Section 1 provides a simple theoretical treatment of bias in initial and central rate estimates, using the conventional means of estimation, viz. adding exposure after exit until the end of the rate year for initial rate estimates. It is immediately seen that $\mathring{m}$ is unbiased for the exponential survival function, and that $\mathring{q}$ is unbiased when the survival function assumes the Balducci form. Extremely concave or convex survival functions are considered. The bias properties of $\mathring{q}$ and $\mathring{m}$ are discussed for the linear survival function, the simple results from which provide the springboard for the methodology of our treatment of concave and convex survival functions. A conclusion terminates the paper.

1. BIAS IN INITIAL AND CENTRAL RATE ESTIMATES

The rate year under investigation is taken as the life year between the xth and $(x+1)$th birthdays.

Let P_{x+t} be the number attaining exact age $x+t$ during the investigation, where $t \varepsilon (0,1)$. The expected number of deaths at age $x+t$, or more precisely between ages $x+t$ and $x+t+dt$ say, is then $\mu_{x+t} \cdot P_{x+t} \cdot dt$. If one were to estimate rates from only this short time interval, one would obtain

$$E(\mathring{m}(t)) = \mu_{x+t} \cdot P_{x+t} \cdot dt / P_{x+t} \cdot dt = \mu_{x+t} \tag{1.1}$$

$$E(\mathring{q}(t)) = \mu_{x+t} \cdot P_{x+t} \cdot dt / [P_{x+t} \cdot dt + \mu_{x+t} \cdot P_{x+t} \cdot dt \cdot (1-t)]$$
$$= \mu_{x+t} / [1 + \mu_{x+t} \cdot (1-t)] \tag{1.2}$$

The estimates $\mathring{m}$ and $\mathring{q}$ have expected values

$$E(\mathring{m}) = \int \mu_{x+t} \cdot P_{x+t} \cdot dt / \int P_{x+t} \cdot dt \tag{1.3}$$

$$E(\mathring{q}) = \int \mu_{x+t} \cdot P_{x+t} \cdot dt / \int [P_{x+t} \cdot dt + \mu_{x+t} \cdot P_{x+t} \cdot dt \cdot (1-t)] \tag{1.4}$$

We have assumed that the exposure at age $x+t$, viz. $P_{x+t} \cdot dt$, is non-random, and that the expectation operator can act separately upon the numerator and denominator of $\mathring{q}$. These assumptions will be adequate provided that the number of decrements is not too large relative to the number of lives in the experience; see Roberts 1986, section 5, for some further justification.

Proportional biases in $\mathring{q}(t)$ and $\mathring{m}(t)$ at each point are defined as

$$pbq(t) = [E(\mathring{q}(t)) - q]/q$$

and

$$pbm(t) = [E(\mathring{m}(t)) - m]/m.$$

We then define a relative proportional bias at each point t:

$$relpb(t) = pbq(t)/pbm(t).$$

The proportional biases in $\mathring{q}$ and $\mathring{m}$ are

$$pbq = [E(\mathring{q}) - q]/q$$

and

$$pbm = [E(\mathring{m}) - m]/m,$$

while the overall relative proportional bias is pbq/pbm.

It is easily seen from equations (1.1) and (1.3) that

$$E(\mathring{m}) \cdot \int P_{x+t} \cdot dt = \int E(\mathring{m}(t)) \cdot P_{x+t} \cdot dt,$$

from which

$$pbm \int P_{x+t} \cdot dt = \int pbm(t) \cdot P_{x+t} \cdot dt,$$

so that the proportional bias in $\mathring{m}$ is a weighted sum of the proportional biases in the estimators at each age. Similarly

$$pbq \int P_{x+t} \cdot dt\, [1 + \mu_{x+t} \cdot (1-t)] = \int pbq(t) \cdot P_{x+t} \cdot dt \cdot [1 + \mu_{x+t} \cdot (1-t)],$$

and the proportional bias in $\mathring{q}$ is a weighted sum of the proportional biases at each age, with weights very close to those used in deriving *pbm* from *pbm*(*t*).

2. EXPONENTIAL AND BALDUCCI CURVES

In general, the expected values $E(\mathring{m})$ and $E(\mathring{q})$ will depend on how P_{x+t} varies with t, which in turn depends on the distribution of movements into and out of the experience. Should the expressions (1.1) or (1.2) turn out not to depend upon t, however, the quantities P_{x+t} have no influence on the expected value, the integrals in (1.3) and (1.4) simply cancelling.

The exponential has a constant force of decrement:

$$1_{x+t} = \exp(-\mu t),$$

where $\mu = -ln(1-q) = m_x$ is the constant force
and $q = q_x$;
and the Balducci curve is defined as (Batten, 1978):

$$1_{x+t} = 1/(1+ht),$$

where $h = q/(1-q)$.
Then $\mu_{x+t} = q/[1-(1-t)q]$,
or equivalently $\mu_{x+t}/[1+\mu_{x+t}(1-t)] = q$.

When the force of mortality is constant, (1.1) is clearly independent of t, and we see that (1.2) is independent of t when the Balducci assumption holds true. Thus $\mathring{m}$ is unbiased when the force of mortality is constant, and when the Balducci assumption holds $\mathring{q}$ is unbiased, regardless of whatever patterns of entrants, new entrants, withdrawals, retirements etc. obtain.

3. THE LINEAR DECREMENTAL CURVE

When the decremental curve is linear,

$$1_{x+t} = 1 - qt, \text{ and } \mu_{x+t} = q/(1-qt).$$

From (1.1) and (1.2) and using the exact relationship $m = q/(1-{\cdot}5q)$, it is easy to show that

$$pbq(t) = 2q(t-{\cdot}5)/(1+q-2qt) \tag{3.1}$$

and

$$pbm(t) = q(t-{\cdot}5)/(1-qt) \tag{3.2}$$

The important point about $pbq(t)$ and $pbm(t)$ is their near linearity in t. This, together with the fact that they practically vanish at $t=\cdot 5$, means the bias arising from exposure at age $x+\cdot 5-t_0$, say, will be balanced by bias from equal exposure at age $x+\cdot 5+t_0$; bias is then approximately zero when exposure is symmetric about $t=\cdot 5$. For practical purposes, exposure should be uniform throughout the rate year for bias to be minimized.

At any point t the ratio of the proportional biases in (3.1) and (3.2) is very close to 2, at least for decremental rates likely to be encountered in practice. The proportional biases in the overall rate estimates are simply the weighted sums (with weights $P_{x+t}\cdot dt$—see section 1) of the corresponding expressions (3.1) and (3.2) over all intervals $(t, t+dt)$, and they can also be expected to be in the ratio of 2:1.*

4. EXTREME CURVATURE IN THE SURVIVAL FUNCTION

For extremely concave or convex survival functions, the proportional biases in q and m estimates should be approximately equal. To see this, consider first a survival function so concave that nearly all the exits occur just before age $x+1$. Then $q \simeq m$.

From formulae (1.3) and (1.4), $\mathring{q} \simeq \mathring{m}$, so that proportional biases in $\mathring{m}$ and $\mathring{q}$ should be nearly equal.

A similar argument obtains for an extremely convex survival function, for which almost all exits occur just after age x. Now $q \simeq m/(1+m)$; but $\mathring{q} \simeq \mathring{m}/(1+\mathring{m})$, and again the proportional biases should be close to each other. Relative proportional bias therefore tends to unity as the curvature becomes infinite.

5. THE QUADRATIC FAMILY OF CURVES

The quadratic family of curves used to model survival functions is defined by

$$1_{x+t} = c - b(t+(a-1)/2)^2 \tag{5.1}$$

where $a>1$ and $b>0$ for concave curves, and $a<-1$ and $b<0$ for convex curves. The parameter a is the principal determinant of curvature: the smaller its absolute value, the greater the curvature, i.e. the more concave or convex the curve. The parameters b and c are fixed by the requirements that 1_{x+t} assume the values 1 and $1-q$ at $t=0$ and 1.

It is easily shown that, for the curve (5.1) as for the linear survival function, the proportional biases in $\mathring{q}(t)$ and $\mathring{m}(t)$ are close to being linear on (0,1), and their ratio is approximately constant. The linearity of $pbq(t)$ and $pbm(t)$ is immediate. Consider $\mu_{x+t} = -1'_{x+t}/1_{x+t}$. The numerator is linear by our choice of survival function, and the denominator close to constant for values of q usually

* This need not be so. Consider quantities y_i and z_i, with each y_i approximately twice the value of the corresponding z_i. Should either Σy_i or Σz_i happen to be close to zero, the ratio $\Sigma y_i/\Sigma z_i$ may well differ substantially from two. The comment nevertheless seems reasonable for practical purposes.

encountered. Thus $E(\mathring{m}(t))$ is nearly linear, and $E(\mathring{q}(t))$ will also be approximately linear from (1.2). We shall see that both $\mathring{m}(t)$ and $\mathring{q}(t)$ are very close to being unbiased at age $t = \cdot 5$, so that the ratio of $pbq(t)$ to $pbm(t)$ will also be roughly constant over (0,1), unless either of these latter quantities is close to zero.

The following values of proportional biases and their ratios at $t = 0$, $\cdot 5$ and 1 are easily found from (1.1) and (1.2):

$$pbq(0) = -(1/a + q - q/a)/(1 + q - q/a)$$
$$pbq(\cdot 5) = -[q/(4a)]/[1 + q/(4a)]$$
$$pbq(1) = (1/a + q)/(1 - q)$$
$$pbm(0) = -[1/a + q/2 - (2q/(3a)) + q/(6a^2)]$$
$$pbm(.5) = -[q/(12a)]/[1 - q/2 + q/(4a)]$$
$$pbm(1) = [1/a + q/2 - q/(3a) + q/(6a^2)]/(1 - q)$$

Recalling that a is no less than 1 in absolute value, first suppose that $|a|$ is small; then $|1/a| \gg q > q/|a|$, and we can ignore q/a. For larger values of $|a|$, q/a is again negligible, in comparison with both q and $1/a$. From the above expressions then, accordingly ignoring all terms in q/a and q/a^2, we obtain:

$$pbq(0) = -(1/a + q)/(1 + q)$$
$$pbq(\cdot 5) = 0$$
$$pbq(1) = (1/a + q)/(1 - q)$$
$$pbm(0) = -(1/a + q/2)$$
$$pbm(\cdot 5) = 0$$
$$pbm(1) = (1/a + q/2)/(1 - q)$$
$$relpb(0) = [(1/a + q)/(1/a + q/2)]/(1 + q)$$
$$relpb(1) = (1/a + q)/(1/a + q/2) = (1/(qa) + 1)/(1/(qa) + \cdot 5) \qquad (5.2)$$

Figure 1 graphs the relative proportional bias (expression (5.2)) against $1/(qa)$. We see that $\mathring{q}$ is unbiased when $a = -1/q$, and $\mathring{m}$ unbiased when $a = -2/q$, these corresponding to approximations to the Balducci and exponential curves respectively. For positive values of a, the ratio of the proportional biases ranges between 1 and 2, depending on the relative values of $1/a$ and q: when a is small, the survival function is very concave and the relative proportional bias close to unity, whereas when a is large, the relative proportional bias tends to 2 as the curve becomes the straight line. This is consistent with the discussions on the linear and extremely concave survival functions above.

As the survival function moves from linearity to the exponential, the bias in $\mathring{m}$ becomes small and the relative proportional bias large; as the survival function becomes more convex than the exponential the relative proportional bias jumps from $+\infty$ to $-\infty$, the bias in $\mathring{m}$ changing sign. The relative proportional bias is zero for the Balducci curve, while for curves more convex than the Balducci, q estimates suffer less bias than m estimates.

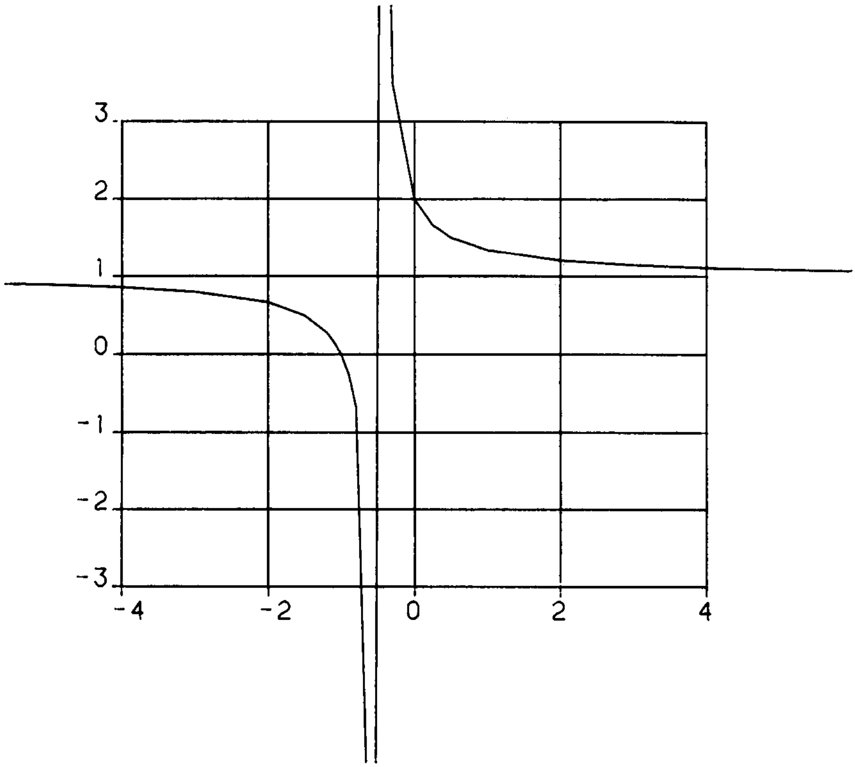

Figure 1. Relative proportional bias against $1/(qa)$.

Four representative standard U.K. life tables (PA(90) males, a(90) males ult., ELT13 males and A67–70 ult.) are concave between the early thirties and the mid-seventies (Roberts, 1986). Comparison of curvatures of these standard tables and the concave quadratic function* indicates that over this age domain appropriate values of a range from about 20 to 60. At age 63 for the A67–70 table, for instance, $q = 2\%$ and an appropriate value of a is about 24, so that the relative bias according to (5.2) is roughly 1.2. For age 52 on the ELT13 table, as a second example, $q = \cdot 01$ and $a = 23$, so that relative proportional bias is about $1 \cdot 1$.

Indications are that the relative proportional bias is close to 1 for ages in the thirties, increasing to about $1 \cdot 2$ by the late fifties, and finally reaching about $1 \cdot 5$ in the seventies. In any case we infer that the relative proportional bias of q and m

* Curvature is defined as $f''/[1 + (f')^2]^{3/2}$, with the prime denoting differentiation. The first and second derivatives were calculated from three consecutive function values, obtained for the standard tables from two successive q values and for the quadratic by taking values at $t = 0$, $\cdot 5$ and 1. The value of a was chosen so that the quadratic had the same curvature as the standard table for the given q value.

estimates for ages between about 30 and 75 can be expected to lie between 1 and 2. At more advanced ages the survival function lies between a straight line and the exponential, so that the relative proportional bias exceeds 2; for the highest ages, in fact, the survival curve is only slightly less convex than the exponential (Roberts, 1986), so that the central rate estimate at these ages will suffer substantially less bias than the initial rate estimate.

The Balducci curve is not often used for modelling survival functions in actuarial work, because the force of decrement falls over the rate year; curves of greater convexity would seem to be even less likely candidates for survival functions, having forces falling still more sharply, and we infer that in general m estimates should be calculated from crude data rather than q estimates.

For given q and a, it is easy to obtain an estimate of the proportional biases in q and m estimates by using equations (1.3) and (1.4). Some algebraic details are provided in an Appendix. Consider first our 63-year-old whose mortality follows the A67–70 table. When exposure to risk is uniform over the rate year, the proportional bias in $\hat{q}$ is about ·02%, so that bias is clearly negligible in comparison with the width of the confidence interval surrounding the estimate. To take another extreme, assume that all exposure is concentrated uniformly in the first half of the rate year; an estimate of q will on average be too low by 3%. Still assuming this latter rather pathologically uneven distribution of exposure, for a total exposure to risk of 10,000 lives the bias is about 1/9 the width of the confidence interval, taking this width as $4\sqrt{(q/E)}$, with E being total exposure; for an exposure to risk of 100,000 the bias is some 1/3 of the width of the confidence interval.

As a final example consider an 87-year-old male with mortality following the ELT13 table. For him the values of a and q are -15 and ·2 respectively, and the magnitude of the mortality rate is such that the assumptions on which our analysis is based become questionable. This point notwithstanding, use of the same procedure, with the same highly skewed distribution of exposure to risk as above, predicts a proportional bias in $\hat{q}$ of -6%; the bias is some 2/3 times the width of the confidence interval for an exposure of 1,000 lives. For a uniform distribution of exposure to risk over the rate year the proportional bias is $+$·5%. The proportional bias in $\hat{q}$ is close to 4 times that in $\hat{m}$ for this individual regardless of the pattern of exposure.

A more detailed discussion of the comparative magnitudes of bias and mean square error, and of the dependence of bias upon the pattern of exposure over the rate year, is provided in a succeeding paper (Roberts, 1987).

CONCLUSION

In estimating m one is in fact assuming that the survival function is an exponential, while for estimates of q the assumed survival function has the Balducci shape. It is hardly surprising that rate estimates are biased when the survival function has not the requisite form, although the generally tiny discrepancy between the expected value of our estimator and the true parameter value is perhaps unexpected.

As well as showing that $\mathring{m}$ is unbiased for m when the survival function is exponential and $\mathring{q}$ unbiased for q for the Balducci survival function, we have shown that the proportional bias in $\mathring{q}$ is double that in $\mathring{m}$ when the survival function is linear, and that the proportional biases in both estimates are approximately the same for extremely concave or convex survival functions. On the assumption that the force of decrement does not fall over the rate year, $\mathring{q}$ will be subject to greater bias than $\mathring{m}$, and one can estimate the extent of the extra bias by calculating curvature of standard tables at similar ages.

Our simple analysis indicates that bias in q and m rate estimates is minimized when exposure is symmetrically distributed around the middle of the rate year. For practical purposes one would merely say that the distribution of exposure should be as uniform as possible over the rate year. Nevertheless, the extent of bias is negligible compared with the standard error of our rate measurements, unless the decremental rate is large, as for mortality rates at advanced ages; or the exposure to risk is very uneven over the rate year; or the exposure to risk is sufficiently large for standard errors of estimates to be small.

It can be argued that our results are specific to the quadratic survival function chosen. Curvatures of standard mortality tables are, however, so low that any smooth function should provide results similar to the above.

More substantial objections are:

(1) Survival functions of greater convexity or concavity may be poorly modelled by this specific curve, chosen purely for analytical convenience.
(2) We have assumed that the constant relative proportional bias at each point of the rate year translates into an overall relative proportional bias of that value (this point was mentioned in passing in the footnote on page 594).

As to the first objection, all one can hope to achieve is to model a survival function by a curve of roughly suitable average curvature over the rate year: the precise functional form should be immaterial, and in any case there would be no way of knowing which family of curves were 'correct', save for investigating decremental rates over smaller age intervals. Some reassurance on the second matter is proved in a more detailed paper by the author (Roberts, 1986), in which simulated investigations to obtain decremental rates, using many patterns of exposure over the rate year and a class of survival functions very different from that used in this paper, gave conclusions consistent with the results reported here.

REFERENCES

BATTEN, R. W. (1978) *Mortality Table Construction.* Prentice-Hall.

ROBERTS, L. A. (1986) *Bias in Decremental Rate Estimates.* OARD Paper, Institute of Actuaries, London.

ROBERTS, L. A. (1987) *Bias and Mean Square Error of Decremental Rate Estimates.* Unpublished.

APPENDIX

We summarize the results of the calculation of $E(\mathring{q})$ when exposure to risk over the rate year assumes a particularly simple form.

Assume that exposure to risk is uniform over (t_0,t_1), and zero elsewhere. From formula (1.4) we see that

$$E(\mathring{q})=\int \mu_{x+t}\cdot dt/[t_1-t_0+\int \mu_{x+t}\cdot dt-\int t\cdot\mu_{x+t}\cdot dt],$$

where

$$\int \mu_{x+t}\cdot dt=ln\ l_0-ln\ l_1,$$

and

$$l_i=c-b[t_i+(a-1)/2]^2.$$

The parameters are given by

$$b=q/a; \qquad c=1+b(a-1)^2/4.$$

Now

$$\int t\cdot\mu_{x+t}\cdot dt=-2(t_1-t_0)-\cdot 5(a-1)\int \mu_{x+t}\cdot dt+\int 2c\cdot dz/(c-bz^2)$$

where $z=t+(a-1)/2$. The value of this last integral, call it K, depends on whether the survival function is convex or concave.

For concave curves $(a>1)$,

$$K=\sqrt{(c/b)}\cdot ln[(1+y_1)(1-y_0)/\{(1-y_1)(1+y_0)\}],$$

where

$$y_i=+\sqrt{(b/c)}\cdot z_i, \qquad \text{and } z_i=t_i+(a-1)/2;$$

for convex curves $(a<-1)$,

$$K=2\sqrt{(-c/b)}\cdot(\arctan y_1-\arctan y_0),$$

where y_i is now $+\sqrt{(-b/c)}\cdot z_i$.