JIA 115 (1988) 1-149

ON GRADUATION BY MATHEMATICAL FORMULA

By D. O. Forfar, B.A., F.F.A., A.I.A., J. J. McCutcheon, M.A., Ph.D., F.F.A. and A. D. Wilkie, M.A., F.F.A., F.I.A.

1. INTRODUCTION

In the course of work undertaken as members of the Executive Committee of the Continuous Mortality Investigation Bureau in the preparation of graduated tables of mortality for the experiences of 1979–82, we have had occasion to make use of and develop a number of statistical techniques with which actuaries may not be familiar, and which are not fully discussed in the current textbook by Benjamin & Pollard (1980), though some of them have been referred to in previous papers by the CMI Committee (1974, 1976). We therefore felt that it would be useful to the profession if we were to present these methods comprehensively in one paper. We do this with the permission of the other members of the CMI Committee, who do not, however, take responsibility for what follows, whether good or bad.

This paper describes theoretical methods. The accompanying report, 'The Graduation of the 1979–82 Experiences', published by the CMI Committee, describes the results of applying these theoretical methods to the actual mortality experiences gathered by the CMI Bureau, and a certain amount of cross-referencing has been inevitable, mostly from the report to this paper.

In Section 2 of the paper we discuss the estimation of various measures of mortality for single ages, based on the data available to the CMI Committee. This is material familiar to actuaries, but discussion of it in a formal way sets the scene for what follows. In Section 3 we briefly comment on the reasons for graduation, and the preliminary calculations that are useful before embarking on a graduation procedure. In Section 4 we describe a family of related formulae that have been found useful for graduation, and which we use frequently in practice. In Section 5 we describe an alternative family, based on splines, that has been found useful in other circumstances. In Section 6 we discuss the various criteria that can be used for finding the 'best' parameters, among them the maximum likelihood and minimum χ^2 parameters. In Section 7 we refer briefly to the numerical methods that can be used to find the optimal parameters for a particular criterion.

In Section 8 we introduce the 'information matrix' from which an estimate of the variance-covariance matrix of the parameters can be found. In Section 9 we discuss the tests that can be applied to a single graduation; some of these are well known, others new. In Section 10 we discuss how one may choose between different graduations, based on different members of one family of formulae. In Section 11 we describe how confidence intervals for the estimated rates of mortality can be obtained, using, in general, simulation methods. In Section 12 we show how to reassess the data and the preferred graduation together. In Section 13 we consider how one may compare two different experiences, both before and after they have been graduated. In Section 14 we show how to construct a complete mortality table on the basis of the graduated rates of mortality, whatever these rates may be $-q_x$, μ_x or m_x.

In Sections 15 to 17 we discuss fully certain examples of these graduation methods as applied to particular experiences chosen from the CMI data. These are particular examples of the data discussed further in the accompanying report.

In our methodology we have attempted to reconcile the traditional methods used by actuaries for many years on both sides of the Atlantic with the more modern work of statisticians in the fields of survival data and life testing. Books describing the latter methods include Mann, Schafer & Singpurwalla (1974), Kalbfleisch & Prentice (1980), Elandt-Johnson & Johnson (1980) and Cox & Oakes (1984). Hogg & Klugman (1984) consider the problem of fitting distributions to data from general insurance; they apply the same sort of methodology to this closely related problem. A recent book in the former line of development is London (1984), which applies statistical methods to Whittaker graduation methods (of which summation formulae are a special case). In our view, graduation by mathematical formula has so many advantages over Whittaker methods that we prefer our method in all cases; at least, we have not come across cases where we have found that a mathematical formula of some kind was not suitable.

A certain amount of mathematical material is presented, for completeness, in the Appendices.

2. ESTIMATION OF MORTALITY RATES FOR SINGLE AGES FROM THE DATA

2.1 *The CMI Data*

We first review the form of data available in investigations undertaken on the lines of the CMI ones. Each investigation carried out by the CMI Bureau is for a particular class of business, for example those purchasing immediate annuities from a life office. It is assumed that the mortality experience of the lives in one investigation is homogeneous, except for variation by age, and possibly by duration. The experience of the two sexes is always kept separate. It has always been found that the experience of the two sexes is very different. Each investigation is carried out for a particular number of calendar years. The results are reported to contributing offices for each calendar year, and they are then grouped into four-yearly periods for fuller reporting in CMI Reports and for possible graduation or preparation of standard tables.

A census of the in-force is carried out at 1 January and 31 December of each calendar year of the investigation, the census for 31 December of one year being identical with that for 1 January of the following year for any one office. Since the number of contributing offices may vary from year to year, the totals for 31 December and 1 January may differ. Within a particular class of business the number of lives or policies (and in some cases also pound amounts) in force is counted, subdivided by age in integral years at the nearest birthday of the life assured, annuitant or pensioner, and (in most investigations) by curtate duration (in years since entry). In some cases offices are not able to classify by nearest age, and use an approximation thereto, which may require that the numbers at neighbouring ages on some other age definition are averaged to give an estimate of the numbers at each nearest age.

During the course of each year the number of deaths by lives, policies or amounts as appropriate is recorded by the contributing offices, subdivided by age nearest birthday at death (and not by an approximation thereto), and (where necessary) also by curtate duration at death.

2.2 *Central and Initial Exposed to Risk*

This form of collection of data allows an estimate to be made of the exact period of exposure to risk of the lives concerned. An exact calculation of this period of exposure would count those days when each life concerned was at risk of entering the count of deaths if he had died on that day, that is, an exact day count. Such a day count is not available, but the periodic censuses can be treated as giving an exact value for the count of the in-force on the day of the census, and the trapezium rule can be used to obtain an approximate integration over the period of investigation. If more frequent censuses were available, say quarterly or monthly, then a more accurate estimate of the exact exposure could be made by using the trapezium rule over the shorter periods between censuses. The trapezium rule simply averages the in-force at the ends of each year (quarter, month) in order to estimate the integral over the period. In some circumstances a more elaborate method such as Simpson's rule might be justifiable, but this has never seemed necessary for the CMI data.

The exposure to risk calculated in this way gives what is called a central exposed to risk. If the period of exposure for each death is continued up to the time when that life would otherwise have left the investigation, and this total extra exposure is added to the central exposed to risk, the initial exposed to risk is obtained. It is usual to estimate this extra period of exposure by adding half the number of relevant deaths, and this is done for the CMI data. In other circumstances a different fraction may be appropriate, or it may be possible to calculate the period of extra exposure exactly. These two forms of exposed to risk are consistent with two different hypotheses about the nature of the mortality process and lead to estimates of alternative types of death-rate—the force of mortality, μ, and the probability of death, q. Both have their justification and their uses.

Division of the number of deaths at a particular age and duration within one investigation by the corresponding central exposed to risk leads to an estimate of μ (or of the central death-rate, m), and division of the number of deaths by the initial exposed to risk leads to an estimate of q, in each case for an appropriate age and duration. We denote the number of deaths recorded in a particular investigation at age x and duration d by $A_{x,d}$, or where duration is not relevant by A_x, or where there is no ambiguity just by A. We denote the corresponding central exposed to risk by $R^c_{x,d}$ or R^c_x, or where there is no ambiguity just by R^c, and the corresponding initial exposed to risk by $R^i_{x,d}$, R^i_x or R^i. In some cases we refer to either R^c or R^i just as R.

We can denote by $\mu^*_{x,d}$ or μ^*_x or just by μ^* the estimate derived as A/R^c, and by $q^*_{x,d}$ or q^*_x or just by q^* the estimate derived as A/R^i. We discuss the justification for these estimators below. The estimators μ^* and q^* can be described as crude rates. It will be explained below that in each case, under certain assumptions, the crude rate would be a satisfactory estimate of the corresponding true rate, if no information about the rates at neighbouring ages were available.

2.3 *Estimation of* q

We first consider the derivation of q^*, the estimator for q. If it is assumed that R persons enter observation at exact age x and exact duration d, and continue under observation until they survive to exact age $x + 1$ and exact duration $d + 1$ or die sooner, that the probability of death within the year for each of them is q, and that the death or survival of each is independent of the death or survival of each of the others, then the appropriate probability model is one of independent Bernoulli trials. The random variable, K, which represents the number of deaths that occur in the year, is binomially distributed with parameters R and q, and the probability of k deaths is

$$P(K = k) = \frac{R!}{k!(R-k)!} q^k (1-q)^{R-k} \qquad (2.3.1)$$

The expected number of deaths is $E[K] = Rq$, and the variance of the number of deaths is $\text{Var}[K] = Rq(1-q)$. Provided the expected number of deaths is reasonably large, in practice bigger than about 5, then K is approximately normally distributed with mean Rq and variance $Rq(1-q)$.

If the value of q is unknown, but it is observed that A of the R persons do die and hence $R - A$ of them survive, then the maximum likelihood estimator of q is $q^* = A/R$. This same estimator is obtained by equating moments (actual = expected). The variance of q^* is $q(1-q)/R$. Deriving this simple result may remind the reader how maximum likelihood estimators are derived.

If the probability of death is q, then the probability that A deaths occur is

$$P(K = A) = \frac{R!}{A!(R-A)!} q^A (1-q)^{R-A} \qquad (2.3.2)$$

The first factor of this expression does not depend on q, so we can describe the likelihood of A deaths as a function of q, $L(q)$, where

$$L(q) = q^A (1-q)^{R-A} \qquad (2.3.3)$$

The natural logarithm of $L(q)$ is

$$L^*(q) = \log L(q) = A \log (q) + (R - A) \log (1 - q). \quad (2.3.4)$$

The value of q that maximizes $L(q)$ is the same as that which maximizes $L^*(q)$. It can be found by differentiating L^* with respect to q and equating to zero the derivative

$$\frac{dL^*(q)}{dq} = \frac{A}{q} - \frac{R - A}{1 - q} \quad (2.3.5)$$

This is zero when $q = A/R$, so that the maximum likelihood estimator of q is

$$q^* = \frac{A}{R} \quad (2.3.6)$$

The expected value of q^* is

$$E[q^*] = E[A/R] = E[A]/R = qR/R = q,$$

and the variance of q^* is

$$\begin{aligned} \text{Var}\,[q^*] &= \text{Var}[A/R] \\ &= \text{Var}[A]/R^2 \\ &= Rq(1 - q)/R^2 \\ &= q(1 - q)/R \end{aligned} \quad (2.3.7)$$

Alternatively the variance of q^* can be derived (at least asymptotically as R increases) from the reciprocal of minus the expected value of the second derivative of $L^*(q)$ with respect to q. This is a special case of a more general result applicable where there are several parameters (see Section 8 below). Minus the second derivative of $L^*(q)$ in this case is

$$-\frac{d^2 L^*(q)}{dq^2} = \frac{A}{q^2} + \frac{R - A}{(1 - q)^2} \quad (2.3.8)$$

The expected value of this, when we put $E[A] = Rq$, is

$$\begin{aligned} E\left[-\frac{d^2 L^*(q)}{dq^2}\right] &= \frac{Rq}{q^2} + \frac{(R - Rq)}{(1 - q)^2} \\ &= \frac{R}{q} + \frac{R}{1 - q} \\ &= \frac{R}{q(1 - q)} \end{aligned} \quad (2.3.9)$$

The reciprocal of this is $q(1 - q)/R$, which indeed equals Var $[q^*]$, as above.

We do not know the true value of q, so in either case we substitute the maximum likelihood estimate, $q^* = A/R$. The variance of q^* can therefore be estimated as

$$\mathrm{Var}[q^*] \simeq \frac{q^*(1 - q^*)}{R} \tag{2.3.10}$$

and from this confidence intervals for q^* can be derived (see § 2.6). Note that the result in this case is exact. In general this derivation is only asymptotically correct, that is the result approaches the correct answer as the sample size increases. Note also that in the above derivation, R represented the initial exposed to risk, R^i.

2.4 *Estimation of* μ

We now consider the derivation of μ^*, the estimator for μ. We assume that a group of persons is observed between ages x and $x + 1$ and durations d and $d + 1$, for various periods within this region, so that person i enters observation at age $x + t_i$ and leaves it either by death or survival at age $x + u_i$, but always within the region $(x, x + 1)$, $(d, d + 1)$. We assume first that the force of mortality μ is constant for all persons within this region, and that the death or survival of each is independent. If the total time that the persons are under observation is

$$R = \sum_i (u_i - t_i) \tag{2.4.1}$$

then the random variable, K, representing the number of deaths that occurs in the period of observation, has a Poisson distribution with mean and variance both equal to $R\mu$ (see Sverdrup, 1965). If the expected number of deaths is reasonably large, in practice bigger than about 5, then K is approximately normally distributed with mean $R\mu$ and variance $R\mu$.

We can derive the likelihood in two ways. At this point we assume that the force of mortality, μ_y, is a function of the attained age y, but not of the duration within the interval $(d, d + 1)$. From the usual actuarial principles we see that the contribution to the likelihood of one person, i, who survives is

$$\exp\left(-\int_{x+t_i}^{x+u_i} \mu_y \mathrm{d}y\right) = {}_{u_i - t_i}p_{x+t_i} \tag{2.4.2}$$

and the contribution to the likelihood of one person, j, who dies is

$$\exp\left(-\int_{x+t_j}^{x+u_j} \mu_y \mathrm{d}y\right)\mu_{x+u_j} = {}_{u_j - t_j}p_{x+t_j} \cdot \mu_{x+u_j} \tag{2.4.3}$$

We now revert to the assumption that μ_x is constant for all ages within the interval $(x, x + 1)$ and equals μ. The contribution to the likelihood of one person, i, who survives is

$$\exp\{-\mu(u_i - t_i)\} \tag{2.4.4}$$

and the contribution to the likelihood of one person, j, who dies is

$$\exp\{-\mu(u_j - t_j)\}\cdot\mu \tag{2.4.5}$$

The total likelihood, $L(\mu)$, is the product of the individual contributions to the likelihood. Thus

$$L(\mu) = \left[\prod_i \exp\{-\mu(u_i - t_i)\}\right]\left[\prod_j \exp\{-\mu(u_j - t_j)\}\mu\right] \tag{2.4.6}$$

where the first term gives the product of the contributions from all those who survive, and the second term gives the product of the contributions of those who die. Therefore

$$\begin{aligned} L(\mu) &= \left[\prod_{\text{all lives}} \exp\{-\mu(u_i - t_i)\}\right]\cdot\left[\prod_{\text{deaths}} \mu\right] \\ &= \exp(-\mu R)\mu^A \end{aligned} \tag{2.4.7}$$

Alternatively, from the fact that K has a Poisson distribution, the probability of A deaths is

$$P(K = A) = \frac{\exp(-\mu R)(\mu R)^A}{A!} \tag{2.4.8}$$

If we ignore terms that do not include μ, we get the likelihood

$$L(\mu) = \exp(-\mu R)\mu^A \tag{2.4.9}$$

as above. The logarithm of the likelihood is

$$L^*(\mu) = -\mu R + A\log(\mu) \tag{2.4.10}$$

and its derivative is

$$\frac{dL^*(\mu)}{d\mu} = -R + \frac{A}{\mu} \tag{2.4.11}$$

which equals zero when

$$\mu = \frac{A}{R} \tag{2.4.12}$$

Hence the maximum likelihood estimator is

$$\mu^* = \frac{A}{R} \tag{2.4.13}$$

The expected value of μ^* is

$$E[\mu^*] = E[A/R] = E[A]/R = \mu R/R = \mu, \tag{2.4.14}$$

and the variance of μ^* is

$$\begin{aligned} \mathrm{Var}[\mu^*] &= \mathrm{Var}[A/R] \\ &= \mathrm{Var}[A]/R^2 \\ &= R\mu/R^2 \\ &= \mu/R \end{aligned} \qquad (2.4.15)$$

Alternatively (at least asymptotically as R increases) the variance of μ^* can be derived (using the same method as before) from the reciprocal of minus the expected value of the second derivative of $L^*(\mu)$ with respect to μ. This is

$$-\frac{\mathrm{d}^2 L^*(\mu)}{\mathrm{d}\mu^2} = \frac{A}{\mu^2} \qquad (2.4.16)$$

The expected value of this, if we put $E[A] = R\mu$, is

$$E\left[-\frac{\mathrm{d}^2 L^*(\mu)}{\mathrm{d}\mu^2}\right] = \frac{R\mu}{\mu^2} = \frac{R}{\mu} \qquad (2.4.17)$$

The reciprocal of this is μ/R which indeed equals $\mathrm{Var}[\mu^*]$. In this case the result is exact. We do not know the true value of μ, so we substitute the maximum likelihood estimate, $\mu^* = A/R$, to get an estimate of the variance of μ^* as

$$\mathrm{Var}[\mu^*] \simeq \frac{\mu^*}{R} \qquad (2.4.18)$$

Note the resemblance of this formula to the variance for q^*, given by equation (2.3.10). In the formula for q^* the exposed to risk, R, is the number of persons who started the year of observation, which corresponds to an initial exposed to risk, R^i. In the formula for μ^* the exposed to risk, R, is the total period of observation lived by those observed, which corresponds to the central exposed to risk, R^c.

2.5 *The Traditional Actuarial Approach*

The circumstances in which these derivations are applicable are wider than we have described above. The traditional actuarial assumption, mentioned briefly by Benjamin & Pollard (1970) and described much more fully by Batten (1978) and Greville (1978), is that the binomial formula for the derivation of q^* remains applicable when the exposure is calculated, not only when a particular number of lives, R, is observed for a full year each (or until previous death), but also when the total exposure is made up from a number of shorter periods of exposure, i.e. the central exposure, to which is added an estimate of the outstanding fraction of a year for the deaths, giving R in total. The Balducci hypothesis that, for $0 \leqslant t \leqslant 1$,

$${}_{1-t}q_{x+t} = (1-t)q_x \qquad (2.5.1)$$

is adduced to justify the procedure. (The Balducci hypothesis is equivalent to the assumption that, for $0 \leqslant t \leqslant 1$, the *reciprocal* of l_{x+t} is a linear function of t.)

This procedure has been used by the CMI Committee in all its previous graduations, explicitly so in its 1974 and 1976 papers. However, Hoem (1980 and 1984) has questioned the validity of $q^* = A/R^i$ as a maximum likelihood estimator in the presence of random withdrawals, and the Balducci assumption, which implies that the force of mortality is falling over each year of age, is an uncomfortable one. Roberts (1987) has pointed out that the possible bias in the estimation of q by this method is almost always larger than that of the estimation of μ on the Poisson model. This possible bias is larger if the force of mortality is changing rapidly over the year of age. Scott (1986) has observed that, if the total exposure is made up from a number of fractional periods of a year, rather than a number of whole years, the variance of the actual number of deaths is increased, and the variance of the estimator q^* also increases. It is very likely that this is the case in practice, as individuals effect policies throughout their years of age and throughout the calendar year.

The Poisson model has been shown by Sverdrup (1965) to hold also for multiple forces of decrement. It is useful also when the underlying force of mortality is a function of two variables, such as age and duration since entry, but can be assumed to be constant over some region $(x, x+1)$, $(d, d+1)$. Scott (1982) has shown that the Poisson model also holds when the force of mortality varies within the year of age, provided that the numbers of lives exposed to risk remain constant (i.e. each person who dies or withdraws is immediately replaced by another person at exactly the same age). In this case the estimator A/R gives an estimate, not of a constant μ, but of

$$\int_x^{x+1} \mu_y \mathrm{d}y \tag{2.5.2}$$

which is not very different from

$$m_x = \frac{\int_x^{x+1} l_y \mu_y \mathrm{d}y}{\int_x^{x+1} l_y \mathrm{d}y} \tag{2.5.3}$$

provided that l does not change rapidly over the year of age $(x, x+1)$.

2.6 *Confidence Intervals*

The true values of q or μ are unknown and q^* or μ^* are the corresponding maximum likelihood estimates. Confidence intervals for these estimates can be obtained in the usual manner, either roughly on the basis of an appropriate normal approximation or accurately with due allowance for the true underlying distribution. If A, the actual number of deaths, is sufficiently large, the simplest normal approximation is generally satisfactory. However, when A is very small or very large (relative to R the exposed to risk) the simplest normal approximation can produce anomalous results—for example, in the case of a confidence interval for q an upper limit which is greater than 1 or a lower limit which is negative, or, in the case of a confidence interval for μ, a negative lower limit. In such a situation it is clearly necessary either to use a better approximation or to derive the confidence interval accurately, having due regard to the true underlying distribution. In the final paragraph of this section we discuss briefly the question as to when it is desirable to work with accurate confidence intervals. For completeness, however, we first describe in greater detail both the approximations to and the derivation of the accurate confidence intervals for q and μ (see Kendall & Stuart, 1979 or Larson, 1982).

Suppose that α is given ($0 < \alpha < \frac{1}{2}$) and that we seek a 100(1–2α) per cent confidence interval for q or μ. In relation to the standard normal distribution let z_α be the abscissa which gives probability α in the upper tail 'above z_α'. (For example, if we wish a 95% confidence interval, we put $\alpha = 0{\cdot}025$ and $z_\alpha = 1{\cdot}960$.)

Let R be the exposed to risk corresponding to A, the actual number of deaths. (Note that R is the initial or central exposure as appropriate.) We consider in turn confidence intervals for (a) q with initial exposures (b) μ and (c) q with central exposures.

(a) *Confidence intervals for q (with initial exposure to risk)*

In this first case (see § 2.3 above) the distribution for A is binomial. Our estimator for q is $q^* = A/R$, which has expected value q. If now we assume that the observed crude rate of mortality is one value from a normal distribution for which the variance is given *exactly* by equation (2.3.10), then the appropriate confidence interval for the mean of the distribution is

$$\left.\begin{aligned} &\left(q^* - z_\alpha\left[\frac{q^*(1-q^*)}{R}\right]^{1/2},\ q^* + z_\alpha\left[\frac{q^*(1-q^*)}{R}\right]^{1/2}\right) \\ \text{i.e.}\quad &\left(\frac{A - z_\alpha[A(1-(A/R))]^{1/2}}{R},\ \frac{A + z_\alpha[A(1-(A/R))]^{1/2}}{R}\right) \end{aligned}\right\} \quad (2.6.1)$$

It should be noted, however, that the upper limit of this interval is greater than 1 when

$$A > \frac{R}{(1 + (z_\alpha^2/R))}$$

and that the lower limit is negative when

$$A < \frac{z_\alpha^2}{(1 + (z_\alpha^2/R))}$$

Crucial to the derivation of the above confidence interval is the assumption of a normal distribution with known variance. Moreover, in general the assumed variance will *not* be the correct value. An alternative approach (used by the CMI Committee, 1974), which avoids any inaccurate assumption concerning the variance and depends solely on a normal approximation, is provided by the fact that, for $z > 0$, the two-sided inequality

$$\frac{2A + z^2 - z[z^2 + 4A(1 - (A/R))]^{1/2}}{2(R + z^2)} < q < \frac{2A + z^2 + z[z^2 + 4A(1 - (A/R))]^{1/2}}{2(R + z^2)}$$

is equivalent to the single inequality

$$\left| \frac{A - Rq}{[Rq(1 - q)]^{1/2}} \right| < z$$

Recall now that the distribution of A has expected value Rq and variance $Rq(1 - q)$. Accordingly, if we assume that A has a normal (rather than binomial) distribution, the appropriate confidence interval for q is

$$\left(\frac{2A + z_\alpha^2 - z_\alpha[z_\alpha^2 + 4A(1 - (A/R))]^{1/2}}{2(R + z_\alpha^2)}, \frac{2A + z_\alpha^2 + z_\alpha[z_\alpha^2 + 4A(1 - (A/R))]^{1/2}}{2(R + z_\alpha^2)} \right) \tag{2.6.2}$$

The lower limit of this interval is never negative and is zero only when $A = 0$. The upper limit of this interval is never greater than 1 and equals 1 only when $A = R$.

Both the intervals (2.6.1) and (2.6.2) are approximate. An accurate confidence interval is readily provided by well-known properties of the binomial distribution. For this interval, if $A = 0$ the lower limit is 0; if $A > 0$, the lower limit is the unique value of q (between 0 and 1) for which

$$\left.\begin{aligned} &\sum_{k=A}^{R} \frac{R!}{k!(R-k)!} q^k (1-q)^{R-k} = \alpha \\ &\text{or, equivalently,} \\ &\sum_{k=0}^{A-1} \frac{R!}{k!(R-k)!} q^k (1-q)^{R-k} = 1 - \alpha \end{aligned}\right\} \tag{2.6.3}$$

If $A = R$, the upper limit of the interval is 1; if $A < R$, the upper limit is the unique root (between 0 and 1) of the equation

$$\left.\begin{aligned} \sum_{k=0}^{A} \frac{R!}{k!(R-k)!} q^k (1-q)^{R-k} &= \alpha \\ \text{or, equivalently} \qquad & \\ \sum_{k=A+1}^{R} \frac{R!}{k!(R-k)!} q^k (1-q)^{R-k} &= 1-\alpha \end{aligned}\right\} \tag{2.6.4}$$

The roots of these equations can readily be found by one of the methods of successive approximation, such as the secant method. (We have found that useful starting values for this method are q^* and $q^* \pm k[q^*(1-q^*)/R]^{1/2}$, where $k = z_\alpha/2$, say.)

(b) *Confidence intervals for μ (with central exposure to risk)*

In this second case (see § 2.4 above) the distribution of A is Poisson. Our estimator for μ is $\mu^* = A/R$, which has expected value μ. If we assume that the observed crude rate of mortality is one value from a normal distribution for which the variance is given *exactly* by equation (2.4.18), then the appropriate confidence interval for μ is

$$\left.\begin{aligned} &\left(\mu^* - z_\alpha\left[\frac{\mu^*}{R}\right]^{1/2},\ \mu^* + z_\alpha\left[\frac{\mu^*}{R}\right]^{1/2}\right) \\ \text{i.e.} \qquad &\left(\frac{A - z_\alpha A^{1/2}}{R},\ \frac{A + z_\alpha A^{1/2}}{R}\right) \end{aligned}\right\} \tag{2.6.5}$$

Note that, if $A < z_\alpha^2$, the lower limit of this interval is negative. As with the corresponding confidence interval for q (i.e. equation (2.6.1) above), this approximation depends on the assumption of a normal distribution with known variance (which will in general be incorrect). To avoid this second assumption we observe that, for $z > 0$, the two-sided inequality

$$\frac{2A + z^2 - z[z^2 + 4A]^{1/2}}{2} < \lambda < \frac{2A + z^2 + z[z^2 + 4A]^{1/2}}{2} \tag{2.6.6}$$

is equivalent to the single inequality

$$\left|\frac{A - \lambda}{\lambda^{1/2}}\right| < z$$

Accordingly, if we assume that A has a normal (rather than Poisson) distribution with mean and variance both equal to $\lambda = R\mu$, the appropriate confidence interval for λ is given by equation (2.6.6) above (with $z = z_\alpha$). Since $\mu = \lambda/R$, the resulting confidence interval for μ is

$$\left(\frac{2A + z_\alpha^2 - z_\alpha[z_\alpha^2 + 4A]^{1/2}}{2R},\ \frac{2A + z_\alpha^2 + z_\alpha[z_\alpha^2 + 4A]^{1/2}}{2R}\right) \tag{2.6.7}$$

The lower limit of this interval is never negative and is zero only when $A = 0$.

An accurate confidence interval for λ is easily found from the true distribution. If $A = 0$, the lower limit is 0; if $A > 0$, the lower limit is the unique positive root of the equation

$$\left.\begin{aligned} \sum_{k=A}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} &= \alpha \\ \text{or equivalently,} \quad \sum_{k=0}^{A-1} e^{-\lambda} \frac{\lambda^k}{k!} &= 1 - \alpha \end{aligned}\right\} \tag{2.6.8}$$

The upper limit is the unique positive root of the equation

$$\left.\begin{aligned} \sum_{k=0}^{A} e^{-\lambda} \frac{\lambda^k}{k!} &= \alpha \\ \text{or equivalently,} \quad \sum_{k=A+1}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} &= 1 - \alpha \end{aligned}\right\} \tag{2.6.9}$$

Substituting $\lambda = R\mu$ in the equations (2.6.8) and (2.6.9) above, we immediately obtain an accurate confidence interval for μ. (Again the secant method may be used to solve these equations.)

(c) *Confidence intervals for q (with central exposure to risk)*

In §6.5 below we show how with central exposures we may proceed directly to a formula graduation of q_x. The Poisson model remains appropriate, the parameter of the distribution being $-R \log(1 - q)$. Accordingly in this situation we can obtain confidence intervals for q by using the results in (b) above and simply replacing μ by $-\log(1 - q)$ i.e. by calculating the value of q corresponding to a lower or upper limit for μ from the equation

$$q = 1 - e^{-\mu}$$

In each of the above cases it is of interest to consider when it is preferable to determine a confidence interval accurately, rather than by a normal approximation. We have investigated the accuracy of the approximate method for different values of α and for various values of R and A (in the case of q) and of A (in the case of μ). We do not report in detail on our investigations. Since the confidence intervals are only a guide, very great accuracy is not required. In general practical rules are as follows: in relation to a confidence interval for q, approximation (2.6.1) will suffice provided that A is not too extreme—say if both A and $R - A$ are greater than 10; in relation to a confidence interval for μ, approximation (2.6.7) will suffice provided that A is not too small—say if A is greater than 10. However, since the relevant calculations are simple for a computer, we have generally used the exact method for $A \leqslant 60$, but this is probably unnecessarily precise in most cases.

3. GRADUATION: THE ESTIMATION OF RATES FOR A NUMBER OF AGES

3.1 *Justification for graduation*

So far we have discussed the estimates of q or μ for only one age. In an actual experience we usually observe data for a number of consecutive ages at integer intervals, so we have two sequences $\{R_x\}$, a sequence of exposed to risk, and $\{A_x\}$, a sequence of numbers of deaths, both described as being 'at age x'. It is usual (and convenient) to index these by an integer variable, x. This is not an essential practice, but its adoption means that the quotient A_x/R_x is often the crude death rate (q, μ or m), not at exact age x ($x + \frac{1}{2}$ for μ), but at some other age $x+b$ ($x+b+\frac{1}{2}$ for μ). In determining b the reader will recognise the problems associated with classifying deaths by age nearest birthday at time of death, age last birthday at the 1st January prior to death, etc. In the case of the CMI studies, in which deaths are always classified by age nearest birthday at time of death, the value of b is $-\frac{1}{2}$. Thus, when one refers to the crude rate of mortality at nearest age x, A_x/R_x^i gives an estimate of $q_{x-1/2}$, and A_x/R_x^c gives an estimate of μ_x, or of $m_{x-1/2}$.

When crude mortality rates for a number of consecutive ages are observed it is usual to find that they appear to run fairly smoothly, at least if the numbers of deaths at each age are largish. The justification for graduating the experience is the assumption that the true mortality rates at each age can be represented by a reasonably simple and smooth mathematical function. Simplicity can be defined as requiring fairly few parameters and smoothness can be defined as having relatively small successive differences when the interval of differencing is taken as one year, as discussed for example by Barnett (1986). Before a mathematical graduation is carried out it is desirable to display the crude data graphically. Since mortality rates typically rise exponentially with age over the main adult age range it is convenient to plot the crude rates with a vertical logarithmic scale having four cycles, e.g. from ·0001 to 1·0. For each age x the confidence interval (q_x^l, q_x^h) described in § 2·6 can be plotted around the crude rate q_x^*, if q is the function under consideration, or (μ_x^l, μ_x^h) can be plotted around the crude rate μ_x^*, if μ is the chosen function. The confidence intervals form gates, within which most of the graduated rates need to fall if the graduation is to be considered satisfactory.

3.2 *Duplicates*

We have assumed so far that there are no duplicates in the investigation, that is, the investigation records either independent lives or policies such that no person holds more than one policy. In practice it is common for individuals to hold more than one policy and hence to appear more than once in the count of exposed to risk or of deaths. This is even more pronounced if the investigation records pound amounts (of sum assured, annuity, etc), since each individual contributes a considerable number of pounds. The adjustments to be made for duplicates will be discussed in § 6·2 below, where it will be shown that the presence of duplicates does not bias the crude rates, but does affect the calculation of their confidence intervals. The effect is that the variance of q^* or μ^*, which enters into the calculation of (q^l, q^h) or (μ^l, μ^h), needs to be multiplied by an appropriate variance ratio, which is at least as big as the average number of policies per life, or pounds per life, as appropriate, and equals this average only if the number of policies or pounds is the same for all lives.

A practical way of allowing for duplicates is to divide the actual deaths and the exposed to risk at each age by the appropriate variance ratio before carrying out any further calculations. As will be shown in § 6·2 the existence of duplicates does not change the expected value of the number of deaths, but does alter the variance. If each value of A_x is replaced by A_x/r_x and each value of R_x is replaced by R_x/r_x, where r_x is the appropriate variance ratio, then most of the calculations can proceed as if there had been no duplicates. (This point is discussed further in § 6·2 below). The appropriate variance ratios may be those based on the actual experience, age by age, or on some other experience that may be considered to have a similar pattern of duplicates. In the latter case, there is no justification for assuming that irregularities in the ratios from age to age are carried forward from one experience to another, and it is therefore desirable to smooth the variance ratios before applying them. For this purpose, however, an elaborate graduation method may not be justified.

4. A FAMILY OF RELATED FORMULAE

4.1 Graduation by mathematical formula is, of course, a well-known technique. In appropriate circumstances the power of modern computers makes it a very valuable tool—even with formulae of some complexity. The celebrated laws due to Gompertz and Makeham, namely

$$\mu_x = Bc^x \tag{4.1.1}$$

and

$$\mu_x = A + Bc^x \tag{4.1.2}$$

are among the earliest examples of formulae adopted for graduation purposes. More recently, the assured lives Table A1967-70, the *a*(90) Table for annuitants, the PA(90) Table for pensioners and the FA1975-78 Table for female assured lives have all been constructed on the basis of graduations by formula. Formulae used include

$$\frac{q_x}{p_x} = A + Hx + Bc^x \tag{4.1.3}$$

(suggested by Barnett—see CMI Committee, 1974) and

$$\frac{q_x}{p_x} = \exp\{\text{pol}(x)\} \tag{4.1.4}$$

(due to Wilkie—see CMI Committee, 1976), where pol(x) denotes an appropriate polynomial in x of low degree, often linear or quadratic.

The right-hand sides of each of the above equations may be considered to be particular examples of a more general expression. Suppose that r and s are non-negative integers, not *both* zero, and that $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{r+s})$ is a vector of 'coefficients'.

Define

$$\text{GM}_{\boldsymbol{\alpha}}^{r,s}(x) = \sum_{i=1}^{r} \alpha_i x^{i-1} + \exp\left\{\sum_{i=r+1}^{r+s} \alpha_i x^{i-r-1}\right\} \tag{4.1.5}$$

with the convention that, if $r = 0$, the right-hand side of equation (4.1.5) is to be interpreted as consisting solely of the exponential term

$$\exp\left\{\sum_{i=1}^{s} \alpha_i x^{i-1}\right\}$$

and that, if $s = 0$, it is to be considered as comprising only the polynomial term

$$\sum_{i=1}^{r} \alpha_i x^{i-1}$$

We adopt this convention throughout this paper. We may also write

$$\text{GM}_{\boldsymbol{\alpha}}^{r,s}(x) = \text{pol}_1(x) + \exp\{\text{pol}_2(x)\},$$

where pol$_1(x)$ and pol$_2(x)$ are polynomials in x of orders r and s respectively. (The order of a polynomial equals 1 plus its degree.) We call $\text{GM}_{\boldsymbol{\alpha}}^{r,s}(x)$ the 'Gompertz–Makeham formula of type (r, s)'—or simply the 'GM(r, s) formula'.

It is trivial to verify that the right-hand sides of equations (4.1.1), (4.1.2), (4.1.3), and (4.1.4) are of the form $\text{GM}_{\boldsymbol{\alpha}}^{r,s}(x)$, with $(r, s) = (0, 2), (1, 2), (2, 2)$ and $(0, n)$—for some positive integer n—respectively.

Note also that, if q_x is defined by equation (4.1.3), then

$$q_x = \frac{\text{GM}_{\boldsymbol{\alpha}}^{2,2}(x)}{1 + \text{GM}_{\boldsymbol{\alpha}}^{2,2}(x)}$$

while equation (4.1.4) implies that

$$q_x = \frac{\text{GM}_{\boldsymbol{\alpha}}^{0,n}(x)}{1 + \text{GM}_{\boldsymbol{\alpha}}^{0,n}(x)}$$

for some positive integer n. Accordingly, for a given coefficient vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{r+s})$ it is convenient to define the 'Logit Gompertz–Makeham formula of type (r, s)'—or simply 'the LGM(r, s) formula'—by the equation

$$\mathrm{LGM}_{\alpha}^{r,s}(x) = \frac{\mathrm{GM}_{\alpha}^{r,s}(x)}{1 + \mathrm{GM}_{\alpha}^{r,s}(x)} \tag{4.1.6}$$

Three types of death rate which arise naturally in different investigations are q_x, μ_x, and m_x. Our remarks above indicate that, with a suitable coefficient vector α, a formula of the type

$$q_x \text{ or } \mu_x \text{ or } m_x = \mathrm{GM}_{\alpha}^{r,s}(x) \text{ or } \mathrm{LGM}_{\alpha}^{r,s}(x)$$

may frequently produce a satisfactory graduation—often with very small values of r and s.

According to circumstances a Gompertz–Makeham (GM) formula or a Logit Gompertz—Makeham (LGM) formula may be the more appropriate. The range of each is of some relevance. The possible range of μ is from zero to infinity, whereas the possible range of q is from zero to 1. Provided that the exponential part (of order s) of a GM formula sufficiently swamps any negative element in the first polynomial (of order r), the possible range of values given by such a formula may be from zero to infinity, so that this formula is potentially more suitable for μ than for q. The possible range of a corresponding LGM formula is from zero to 1, so that this formula may be more suitable for q than for μ. These are not, however, absolute rules.

4.2 *Orthogonal polynomials*

In its general form the GM(r, s) formula contains two polynomials—(only one, if either r or s is zero). In equation (4.1.5) above these polynomials are defined (by the coefficient vector $\boldsymbol{\alpha}$) naturally in terms of successive powers of x, i.e. $\{1, x, x^2, x^3, \ldots\}$. This might be considered as the 'obvious' basis for polynomials. In certain situations, however, it may be more appropriate to use an alternative basis. Any sequence of polynomials $\{p_0(x), p_1(x), p_2(x), \ldots\}$, in which $p_i(x)$ $(i = 0, 1, 2, \ldots)$ is of degree i, provides a basis for the set of polynomials.

For example, consider the first type of Chebycheff polynomials $\{C_i(x)\}$, defined by the equations

$$\left.\begin{aligned} C_0(x) &= 1 \\ C_1(x) &= x \end{aligned}\right\} \tag{4.2.1}$$

and the recurrence relation

$$C_{n+1}(x) = 2xC_n(x) - C_{n-1}(x) \quad (n \geqslant 1) \tag{4.2.2}$$

Thus

$$C_2(x) = 2x^2 - 1, \quad C_3(x) = 4x^3 - 3x, \quad C_4(x) = 8x^4 - 8x^2 + 1$$

etc. (see Conte & de Boor, 1980). If $p(x)$ is *any* given polynomial of order n (i.e. of degree $n - 1$), then $p(x)$ is *uniquely* expressible in the form

$$p(x) = \sum_{i=1}^{n} \alpha_i C_{i-1}(x)$$

(For example, if $n = 4$ and $p(x) = 1 + x + x^2 + x^3$, then $\alpha = [\frac{3}{2}, \frac{7}{4}, \frac{1}{2}, \frac{1}{4}]$.) It is this uniqueness property which characterises a basis.

The idea of an *orthogonal* basis arises naturally in the problem of fitting a curve to a given set of data. Given m 'data' points $\{(x_j, y_j), j = 1, \ldots, m\}$ and a set of positive weights $\{w_1, \ldots, w_m\}$ we may for example try to find $f_n(x)$, the polynomial of order n which best fits the data, in the sense that the value of

$$\sum_{j=1}^{m} w_j[y_j - f_n(x_j)]^2 \tag{4.2.3}$$

is minimised.

If the 'natural' basis $\{1, x, x^2, \ldots x^{n-1}\}$ is used to define $f_n(x)$, we write

$$f_n(x) = \sum_{i=1}^{n} \alpha_i x^{i-1}$$

Similarly we may also find $f_{n+1}(x)$, the best-fitting polynomial of order $n + 1$. This will be

$$f_{n+1}(x) = \sum_{i=1}^{n+1} \alpha'_i x^{i-1}$$

Note that in the last equation the coefficient vector is $\boldsymbol{\alpha}' = (\alpha'_1, \alpha'_2, \ldots, \alpha'_n, \alpha'_{n+1})$, whereas in the previous equation it is $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$. In general the values of $\alpha'_1, \alpha'_2, \ldots \alpha'_n$ may be very different from the values of $\alpha_1, \alpha_2, \ldots, \alpha_n$. Thus, when the natural basis is used, the consequence of increasing by one the degree of the best-fitting polynomial is therefore not immediately obvious from consideration of the value of the 'additional' coefficient α'_{n+1} alone.

It is, however, possible to choose an 'orthogonal' basis (which in this particular situation will depend on $\{(x_i, w_i), i = 1, \ldots, m\}$ such that the best-fitting polynomial of order $n + 1$ is obtained from the best-fitting polynomial of order n simply by adding a multiple (α_{n+1}, say) of the $(n + 1)^{\text{th}}$ basis function. In this situation the magnitude of the additional coefficient may well be capable of a simple interpretation and it is perhaps of interest to describe briefly the appropriate orthogonal basis.

For a given set of data points $\{(x_j), j = 1, \ldots, m\}$ and positive weights $\{w_1, \ldots, w_m\}$ the basis of orthogonal polynomials, $\{p_i(x), i = 0, 1, 2, \ldots\}$ (with $p_i(x)$ of degree i), must be chosen so that

$$\sum_{j=1}^{m} w_j p_r(x_j) p_s(x_j) = \begin{cases} 0 & \text{if } r \neq s \\ \mathrm{e}_r & \text{if } r = s \end{cases} \tag{4.2.4}$$

where e_r is some specified non-zero real number (usually 1). (It is readily verified that equation (4.2.4) does indeed define a unique set of polynomials (see, for example, Conte & de Boor, 1980).)

Having constructed our sequence of orthogonal polynomials, we define $f_n(x)$, the best-fitting polynomial of order n, as

$$f_n(x) = \sum_{i=1}^{n} \alpha_i p_{i-1}(x) \tag{4.2.5}$$

the coefficients $\{\alpha_i\}$ being chosen to minimise the expression (4.2.3) above. This means that the coefficients must be chosen to minimize

$$S = \sum_{j=1}^{m} w_j[y_j - \sum_{i=1}^{n} \alpha_i p_{i-1}(x_j)]^2 \tag{4.2.6}$$

The appropriate coefficient vector $\boldsymbol{\alpha}$ is easily obtained by noting that, at the minimum point, all the partial derivatives $(\partial S/\partial \alpha_k)$ $(k = 1, \ldots, n)$ are zero. For $k = 1, \ldots, n$ we therefore have

$$\sum_{j=1}^{m} w_j \cdot 2[y_j - \sum_{i=1}^{n} \alpha_i p_{i-1}(x_j)] \cdot (-p_{k-1}(x_j)) = 0$$

i.e.

$$\sum_{j=1}^{m} w_j y_j p_{k-1}(x_j) = \sum_{j=1}^{m} w_j \sum_{i=1}^{n} \alpha_i p_{i-1}(x_j) p_{k-1}(x_j)$$

Interchanging the order of summation in the right-hand side of this last equation, we obtain

$$\sum_{j=1}^{m} w_j y_j p_{k-1}(x_j) = \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{m} w_j p_{i-1}(x_j) p_{k-1}(x_j) = \alpha_k \cdot \sum_{j=1}^{m} w_j [p_{k-1}(x_j)]^2$$

by equation (4.2.4) above.

Hence

$$\alpha_k = \frac{\sum_{j=1}^{m} w_j y_j p_{k-1}(x_j)}{\sum_{j=1}^{m} w_j [p_{k-1}(x_j)]^2} \tag{4.2.7}$$

which does *not* depend on n. This means that, if $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ defines the best-fitting polynomial of order n and $\boldsymbol{\alpha}' = (\alpha'_1, \alpha'_2, \ldots, \alpha'_{n+1})$ defines the best-fitting polynomial of order $n + 1$, then

$$\alpha'_1 = \alpha_1, \quad \alpha'_2 = \alpha_2, \ldots, \alpha'_n = \alpha_n$$

The value of the 'additional' coefficient α'_{n+1} may be capable of a relatively simple interpretation.

The 'continuous' form of orthogonal polynomials, in relation to a particular real interval $[a, b]$ and given positive 'weight function' $w(x)$, is obtained by constructing a sequence of polynomials $\{p_i(x), i = 0, 1, 2, \ldots\}$ such that

$$\int_a^b w(x)p_r(x)p_s(x)\mathrm{d}x = \begin{cases} 0 & \text{if } r \neq s \\ \mathrm{e}_r & \text{if } r = s \end{cases} \qquad (4.2.8)$$

where, as before, e_r is some non-zero real number. (This corresponds to equation (4.2.4) above.)

In relation to the interval $[-1, 1]$ the Chebycheff polynomials described above are orthogonal with weight function $w(x) = [1 - x^2]^{-1/2}$. In relation to the same interval when $w(x) \equiv 1$, we obtain the Legendre polynomials $\{L_i(x)\}$, defined by the initial values

$$\left.\begin{aligned} L_0(x) &= 1 \\ L_1(x) &= x \end{aligned}\right\} \qquad (4.2.9)$$

and the recurrence relation

$$(n + 1)L_{n+1}(x) = (2n + 1)xL_n(x) - nL_{n-1}(x) \quad (n \geqslant 1) \qquad (4.2.10)$$

There exist other well-known sets of polynomials orthogonal, each for a specified weight function $w(x)$, over the interval $[-1, 1]$.

4.3 *Age scaling*

It is sometimes convenient to express our Gompertz–Makeham or Logit Gompertz–Makeham (r, s) formula in terms of a particular set of orthogonal polynomials. In this case some form of age scaling is usually required. The polynomials are expressed in powers not of x, but of $t = (x - u)/v$, where u and v are suitably chosen. We may choose u and v so that over the range of ages in question the range of t is 0 to 1 or -1 to 1, or at least approximately so. For example, in a pensioners' investigation over the age-range 60 to 110 (say) we might let $t = (x - 85)/25$. Age 60 then corresponds to $t = -1$ and age 110 to $t = 1$.

We can bring together the above ideas as follows. Suppose that $\{p_i(x), i = 0, 1, 2, \ldots\}$ is a given basis of polynomials and that u and v are specified. For a given coefficient vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{r+s})$ we define

$${}^{\{p\}}_{u,v}\mathrm{GM}^{r,s}_{\boldsymbol{\alpha}}(x) = \sum_{i=1}^{r} \alpha_i p_{i-1}\left(\frac{x-u}{v}\right) + \exp\left\{\sum_{i=r+1}^{r+s} \alpha_i p_{i-r-1}\left(\frac{x-u}{v}\right)\right\} \qquad (4.3.1)$$

where, as before, the convention is adopted that, if $r = 0$, the first term on the right-hand side of the equation is omitted and that the second term is omitted when $s = 0$.

Following the discussion in §4.1 above, we may also define

$$ {}^{\{p\}}_{u,v}\mathrm{LGM}^{r,s}_{\alpha}(x) = \frac{{}^{\{p\}}_{u,v}\mathrm{GM}^{r,s}_{\alpha}(x)}{1 + {}^{\{p\}}_{u,v}\mathrm{GM}^{r,s}_{\alpha}(x)} \tag{4.3.2} $$

To the casual reader the notation of these last two equations may appear somewhat daunting! Although the full notation is necessary for total precision, fortunately it is seldom necessary to use it completely. Provided that the set of basis polynomials $\{p_i(x)\}$ and the values of u and v are clearly understood, we may revert to our earlier notation. (Such would be the situation, for example, if we use the Chebycheff polynomials described above and the variable $t = (x - 70)/50$, as in § 15.)

In this case the left-hand sides of equations (4.3.1) and (4.3.2) may be replaced by the simpler notations $\mathrm{GM}^{r,s}_{\alpha}(x)$ and $\mathrm{LGM}^{r,s}_{\alpha}(x)$ respectively. Further simplification may be possible. For example, if we are discussing the class of formula for which $r = 1$ and $s = 2$ and the only item not clearly specified is the coefficient vector α, we may use the shorter notation $\mathrm{GM}_{\alpha}(x)$ and $\mathrm{LGM}_{\alpha}(x)$ respectively.

5. SPLINES AS AN ALTERNATIVE FORMULA

5.1 In the previous section we have restricted our discussion to the classes of GM(r, s) and LGM(r, s) formulae. In § 6 below we indicate how these formulae may be used in curve-fitting. It is important, however, to realise that the remarks in § 6 apply in a completely general context and not simply in relation to these two classes of formulae. In certain situations different formulae may be appropriate. For example, one might consider an equation of the kind

$$ q_x \text{ or } \mu_x \text{ or } m_x = f_{\alpha}(x) $$

with

$$ f_{\alpha}(x) = \frac{\alpha_1 + \alpha_2\alpha_3^x}{1 + \alpha_4\alpha_3^x} \tag{5.1.1} $$

(as used in the a(55) mortality table for annuitants) or

$$ f_{\alpha}(x) = \alpha_1 + \alpha_2\alpha_3^x[\alpha_4\alpha_3^{-2x} + 1 + \alpha_5\alpha_3^x]^{-1} \tag{5.1.2} $$

(as used in the A1949-52 mortality table for assured lives). Other formulae have proved useful in practical situations (see, for example, Heligman & Pollard, 1980).

An alternative tool is provided by the use of *splines*. These have formed the basis of graduation for the last two sets of national life tables and can be useful in a wide variety of practical situations (see McCutcheon, 1981, 1984, and 1987). Splines are just as much mathematical formulae as any of the examples described above. Since, however, actuaries may be less familiar with the use of splines, in our next paragraph we describe very briefly their salient features. For a more detailed discussion the reader should refer to de Boor, (1978).

5.2 *Splines*

A spline is a *piecewise* polynomial function for which the maximum possible number of derivatives exist. More precisely, suppose that $a = x_0 < x_1 < \ldots < x_{m+1} = b$. A spline s of degree k, defined on the interval $[a, b]$ with internal "knots" $x_1, x_2, \ldots, x_m$, is a function such that on each subinterval $[x_i, x_{i+1}]$ $(i = 0, 1, \ldots, m)$ $s(x)$ is given by a polynomial in x of degree k (or less). Moreover, the $(m + 1)$ polynomials which define s fit together in such a manner that s is differentiable $(k - 1)$ times at each of the internal knots.

It is this last condition which distinguishes splines from other piecewise polynomial functions.

Let $n = m + k + 1$. It is easily seen that, for specified internal knot positions $\{x_1, x_2, \ldots, x_m\}$, n 'coefficients' (or parameters) are needed to define a spline s of degree k (see McCutcheon, 1981). (The spline is simply a linear combination of appropriate 'basis' functions.) We may therefore write

$$s(x) = f_{\alpha}(x) \tag{5.2.1}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ is the vector of coefficients defining the spline. In the above discussion, which relates to a spline with predetermined knot positions, $f_{\boldsymbol{\alpha}}(x)$ is *linear* in each of the coefficients $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. By considering the knot positions themselves to be further 'coefficients', we can define the most general m-knot spline of degree k on the interval $[a, b]$. In this case we let $n = m + (m + k + 1) = 2m + k + 1$ and put

$$s(x) = g_{\boldsymbol{\alpha}}(x) \tag{5.2.2}$$

where, as before, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ is the defining coefficient vector. Now, however, the coefficients $\alpha_1, \alpha_2, \ldots, \alpha_m$ must all lie in the interval (a, b) and specify the positions of the knots, while $\alpha_{m+1}, \alpha_{m+2}, \ldots, \alpha_n$ then determine the spline in relation to these knots (duly ordered) by an appropriate basis. Multiple knots may occur.

This latter situation gives rise to the technique of 'variable-knot' spline graduation, described by McCutcheon (1984). It is important to realise that, although in this case $g_{\boldsymbol{\alpha}}(x)$ is non-linear in $\alpha_1, \alpha_2, \ldots, \alpha_m$ (the knot positions), there is no difference in principle between more traditional formulae (such as those described above) and splines. In all cases the curve-fitting formula is defined by an appropriate coefficient vector $\boldsymbol{\alpha}$.

In practice cubic splines (for which $k = 3$) are often a useful tool.

6. CRITERIA FOR OPTIMISATION

6.1 In this section we consider alternative methods of obtaining graduations by formula for q-type, μ-type, and m-type crude rates of mortality. The principal methods which we describe in some detail are (1) *maximum likelihood* and (ii) *minimum* χ^2. One advantage of maximum likelihood methods is that they lead naturally to estimates for the variance–covariance matrix of the underlying parameters—i.e. the coefficients of the particular formula which we are using. (See § 8 below.) As the volume of data increases, the distribution of maximum likelihood estimators tends to the normal distribution. Asymptotically such estimators are unbiased and of minimum variance.

If computing power is limited, or if one is dependent on certain of the available packages, it may be simpler to graduate by the minimum χ^2 method. In this case, having obtained the coefficients of the best graduation, by falling back on certain asymptotic properties, we may often estimate the appropriate variance–covariance matrix quite simply by considering the likelihood function —without the need to maximize it. This is discussed in § 8 below.

It is worth noting that, by making reasonable approximations, in all cases with sufficient data we can reduce the maximization of the likelihood to an equivalent minimum χ^2 calculation. This is discussed below.

For graduation of each of q, μ, and m, we consider first the appropriate likelihood and χ^2 functions. In § 7 we refer briefly to possible ways of carrying out the necessary calculations.

For reference purposes in the Appendices we give further details of some of the calculations which may be required, such as the calculation of certain first and second partial derivatives.

In most practical situations, maximum likelihood and minimum χ^2 lead to very similar graduations, although the resulting coefficient vectors are generally different for the two solutions. This is illustrated by examples later in this paper (see § 15 below).

Suppose, then, that we are given an appropriate function $f_{\boldsymbol{\alpha}}(x)$ with n coefficients $(\alpha_1, \alpha_2, \ldots, \alpha_n)$. We discuss below how this function can be used to graduate the three kinds of mortality rate. (In the practical examples in § 15 below we shall let

$$f_{\boldsymbol{\alpha}}(x) \;=\; \mathrm{GM}_{\boldsymbol{\alpha}}^{r,s}(x)$$

or

$$f_{\boldsymbol{\alpha}}(x) \;=\; \mathrm{LGM}_{\boldsymbol{\alpha}}^{r,s}(x)$$

where $r + s = n$ and r and s are suitably chosen. At present, however, it is sufficient to consider a completely general situation.)

6.2 *Graduation of* q-*type rates (using initial exposures)*

Let

$$q_x \;=\; q_x(\boldsymbol{\alpha}) \;=\; f_{\boldsymbol{\alpha}}(x) \qquad (6.2.1)$$

The notation q_x or $q_x(\alpha)$ thus denotes the rate of mortality at exact age x. The latter form, $q_x(\boldsymbol{\alpha})$, serves simply to remind us that q_x depends on $\boldsymbol{\alpha}$.

We assume first that there are no duplicates and that the available exposures $\{R_x\}$ are in initial form. In this simplest situation a binomial model usually provides the basis for further analysis, independence of the data being assumed at distinct ages. We therefore assume that

$$A_x \sim B(R_x, q_{x+b}(\boldsymbol{\alpha})) \tag{6.2.2}$$

the notation $B(N, q)$ being used to denote a binomial distribution with parameters N and q. (It should, however, be noted that even for this simple model some form of approximation may be implicit in our assumptions, as discussed in § 2.5.)

(a) *Maximum likelihood*
If the values of R_x and A_x are known for $n \leqslant x \leqslant m$ and A_x/R_x is the crude rate of mortality at exact age $x + b$, the likelihood function (when there are no duplicates) is

$$L(\boldsymbol{\alpha}) = \prod_{x=n}^{m} [q_{x+b}(\boldsymbol{\alpha})]^{A_x}[1 - q_{x+b}(\boldsymbol{\alpha})]^{R_x - A_x} \tag{6.2.3}$$

Maximizing L is equivalent to maximizing the logarithm of L. Accordingly we consider

$$\begin{aligned} L_1(\boldsymbol{\alpha}) &= \log L(\boldsymbol{\alpha}) \\ &= \sum_{x=n}^{m} \{A_x \log q_{x+b}(\boldsymbol{\alpha}) + (R_x - A_x)\log(1 - q_{x+b}(\boldsymbol{\alpha}))\} \end{aligned} \tag{6.2.4}$$

The value of $\boldsymbol{\alpha}$ is determined to maximize $L_1(\boldsymbol{\alpha})$. In paragraph (c) below we describe how allowance may be made for duplicates.

(b) *Minimum* χ^2
In the absence of duplicates the number of deaths 'at age x' has mean $R_x q_{x+b}(\boldsymbol{\alpha})$ and variance $R_x q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})]$. In this situation we define the 'relative deviation' at age x to be

$$z_x = [A_x - R_x q_{x+b}(\alpha)]/[R_x q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})]]^{1/2} \tag{6.2.5}$$

and let

$$\chi^2(\boldsymbol{\alpha}) = \sum_{x=n}^{m} z_x^2 = \sum_{x=n}^{m} \frac{[A_x - R_x q_{x+b}(\boldsymbol{\alpha})]^2}{R_x q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})]} \tag{6.2.6}$$

In calculating $\chi^2(\boldsymbol{\alpha})$ for an investigation based on policies allowance must be made for duplicates. This is easily done by defining the 'variance ratio' at age x to be

$$r_x = \sum_i i^2 f_x^i \Big/ \sum_i i f_x^i \tag{6.2.7}$$

where f_x^i ($i = 1, 2, \ldots$) denotes the proportion of policyholders at age x who have i policies. (Thus $\sum_i f_x^i = 1$). When there are duplicates the variance of the number of claims at age x is increased by the factor r_x, so that the appropriate definition of $\chi^2(\boldsymbol{\alpha})$ becomes

$$\chi^2(\boldsymbol{\alpha}) = \sum_{x=n}^{m} \frac{[A_x - R_x q_{x+b}(\boldsymbol{\alpha})]^2}{r_x R_x q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})]} \tag{6.2.8}$$

We then choose $\boldsymbol{\alpha}$ to minimize $\chi^2(\boldsymbol{\alpha})$. (See Seal (1943), Daw (1946), Beard & Perks (1949), Daw (1951) and CMI Committee (1957) for the derivation of this result and for further discussion of the effect of duplicates in a mortality experience.)

The following remarks relating to the treatment of duplicates are perhaps of interest.

Let A be a random variable representing the number of deaths at age x (so that the observed value of A is A_x). We have already commented that in the absence of duplicates

$$E[A] = R_x q_{x+b}(\boldsymbol{\alpha}) \tag{6.2.9}$$

and

$$\mathrm{Var}[A] = R_x q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})] \tag{6.2.10}$$

while, if there are duplicates, the value of $E[A]$ is unchanged but

$$\mathrm{Var}[A] = r_x R_x q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})] \tag{6.2.11}$$

In this latter case let

$$A' = \frac{A}{r_x} \tag{6.2.12}$$

and

$$R'_x = \frac{R_x}{r_x} \tag{6.2.13}$$

Then

$$E[A'] = \frac{1}{r_x} E[A] = \frac{R_x}{r_x} q_{x+b}(\boldsymbol{\alpha}) = R'_x q_{x+b}(\boldsymbol{\alpha}) \tag{6.2.14}$$

and

$$\mathrm{Var}[A'] = \frac{1}{r_x^2} \mathrm{Var}[A] = \frac{R_x}{r_x} q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})]$$

$$= R'_x q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})] \tag{6.2.15}$$

These last two equations are in precisely the same form as equations (6.2.9) and (6.2.10) above, with A' and R'_x replacing A and R_x respectively. If, therefore, before proceeding to the graduation we divide *both* the number of actual deaths and the exposure at each age by the appropriate variance ratio, we obtain an empirical procedure whereby the presence of duplicates may be ignored.

At this stage it is of interest to note that in practice the alternative procedures of maximizing the likelihood and minimizing χ^2 generally produce very similar graduations. This can be expected on theoretical grounds, as was observed by N.L. Johnson in his remarks in the discussion of Barnett (1951). One reason for this is illustrated by the following discussion (see Van der Waerden, 1969).

Note first that, if t is small, then

$$\log(1 - t) + t \simeq -\tfrac{1}{2}t^2 \qquad (6.2.16)$$

We consider the terms in the log-likelihood $L_1(\boldsymbol{\alpha})$ (equation 6.2.4 above). In the discussion below c_x, d_x, e_x etc. denote expressions which, although depending on A_x and R_x, do *not* depend on $q_{x+b} = q_{x+b}(\boldsymbol{\alpha})$.

Now

$$\log q_{x+b} = \log\left(\frac{R_x q_{x+b}}{A_x}\right) + \log A_x - \log R_x$$

Therefore

$$\begin{aligned}
A_x \log q_{x+b} &= A_x \log\left(\frac{R_x q_{x+b}}{A_x}\right) + c_x \\
&= A_x\left\{\log\left[1 - \frac{A_x - R_x q_{x+b}}{A_x}\right]\right\} + c_x \\
&= A_x\left\{\log\left[1 - \frac{A_x - R_x q_{x+b}}{A_x}\right]\right. \\
&\qquad \left. + \frac{A_x - R_x q_{x+b}}{A_x}\right\} - A_x + R_x q_{x+b} + c_x \\
&\simeq -\frac{1}{2} A_x \left(\frac{A_x - R_x q_{x+b}}{A_x}\right)^2 + R_x q_{x+b} + d_x
\end{aligned}$$

by (6.2.16) above, if $R_x q_{x+b} \simeq A_x$.

Thus for an acceptable graduation

$$A_x \log q_{x+b} \simeq -\frac{1}{2A_x}(A_x - R_x q_{x+b})^2 + R_x q_{x+b} + d_x \qquad (6.2.17)$$

Similarly

$$\begin{aligned}
\log(1 - q_{x+b}) &= \log\left(\frac{R_x(1 - q_{x+b})}{R_x - A_x}\right) - \log R_x \\
&\qquad + \log(R_x - A_x)
\end{aligned}$$

Therefore

$$
\begin{aligned}
&(R_x - A_x)\log(1 - q_{x+b}) \\
&= (R_x - A_x)\log\left(\frac{R_x(1 - q_{x+b})}{R_x - A_x}\right) + e_x \\
&= (R_x - A_x)\log\left[1 - \frac{R_x q_{x+b} - A_x}{R_x - A_x}\right] + e_x \\
&= (R_x - A_x)\left\{\log\left[1 - \frac{R_x q_{x+b} - A_x}{R_x - A_x}\right]\right. \\
&\quad \left. + \frac{R_x q_{x+b} - A_x}{R_x - A_x}\right\} - R_x q_{x+b} + A_x + e_x \\
&\simeq -\frac{1}{2}(R_x - A_x)\left(\frac{R_x q_{x+b} - A_x}{R_x - A_x}\right)^2 - R_x q_{x+b} + f_x,
\end{aligned}
$$

if $R_x q_{x+b} \simeq A_x$ (again by equation (6.2.16) above).

Thus, for an acceptable graduation,

$$
\begin{aligned}
&(R_x - A_x)\log(1 - q_{x+b}) \\
&\simeq -\frac{1}{2(R_x - A_x)}(A_x - R_x q_{x+b})^2 - R_x q_{x+b} + f_x
\end{aligned}
\qquad (6.2.18)
$$

By adding equations (6.2.17) and (6.2.18), we obtain (since the terms $R_x q_{x+b}$ cancel each other)

$$
\begin{aligned}
&A_x \log q_{x+b} + (R_x - A_x)\log(1 - q_{x+b}) \\
&\simeq -\frac{1}{2}(A_x - R_x q_{x+b})^2\left(\frac{1}{A_x} + \frac{1}{R_x - A_x}\right) + g_x \\
&= -\frac{1}{2}(A_x - R_x q_{x+b})^2 \frac{R_x}{A_x(R_x - A_x)} + g_x \\
&= -\frac{1}{2}\frac{(A_x - R_x q_{x+b})^2}{R_x \frac{A_x}{R_x}\left(1 - \frac{A_x}{R_x}\right)} + g_x \\
&\simeq -\frac{1}{2}\frac{(A_x - R_x q_{x+b})^2}{R_x q_{x+b}(1 - q_{x+b})} + g_x
\end{aligned}
\qquad (6.2.19)
$$

if $A_x \simeq R_x q_{x+b}$.

Hence, summing over all ages, we obtain

$$
L_1(\alpha) = \sum_x \{A_x \log q_{x+b} + (R_x - A_x)\log(1 - q_{x+b})\}
$$

$$\simeq -\frac{1}{2}\sum_x \left\{\frac{(A_x - R_x q_{x+b})^2}{R_x q_{x+b}(1 - q_{x+b})} + g_x\right\}$$

(by equation (6.2.19))

$$= -\frac{1}{2}\chi^2(\boldsymbol{\alpha}) + h \qquad (6.2.20)$$

where $h = -\frac{1}{2}\Sigma_x g_x$ does *not* depend on $\boldsymbol{\alpha}$. Thus we see that maximizing $L_1(\boldsymbol{\alpha})$ is equivalent to maximizing $-\frac{1}{2}\chi^2(\boldsymbol{\alpha})$, which itself is equivalent to *minimizing* $\chi^2(\boldsymbol{\alpha})$.

(c) *Maximum likelihood approximations and allowance for duplicates*
Recall that, when there are duplicates, the distribution for A_x has mean $R_x q_{x+b}(\boldsymbol{\alpha})$ and variance $r_x R_x q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})]$, where r_x is given by equation (6.2.7) above. (When there are no duplicates, $r_x = 1$.) If we use the normal approximation for the distribution function for A_x, the likelihood function becomes

$$L(\boldsymbol{\alpha}) = \prod_{x=n}^{m} \frac{1}{[2\pi r_x R_x q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})]]^{1/2}} \times \exp\left\{\frac{-[A_x - R_x q_{x+b}(\boldsymbol{\alpha})]^2}{2 r_x R_x q_{x+b}(\boldsymbol{\alpha})[(1 - q_{x+b}(\boldsymbol{\alpha}))]}\right\} \qquad (6.2.21)$$

This last equation implies that

$$\log L(\boldsymbol{\alpha}) = -\frac{1}{2}\sum_{x=n}^{m}\left\{\log(2\pi r_x R_x q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})]) + \frac{[A_x - R_x q_{x+b}(\boldsymbol{\alpha})]^2}{r_x R_x q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})]}\right\}$$

Omitting the terms which do not depend on $\boldsymbol{\alpha}$, we see that the function to be maximized is

$$L_2(\boldsymbol{\alpha}) = -\frac{1}{2}\sum_{x=n}^{m}\left\{\log(q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})]) + \frac{[A_x - R_x q_{x+b}(\boldsymbol{\alpha})]^2}{r_x R_x q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})]}\right\} \qquad (6.2.22)$$

Empirical investigations indicate that in many situations the logarithmic term on the right-hand side of this last equation is much less sensitive to changes in the vector $\boldsymbol{\alpha}$ than the other term—at least in the neighbourhood of the optimal solution. This means that a good approximation to the optimal solution can often be obtained by omitting the logarithmic term and maximising

$$L_3(\boldsymbol{\alpha}) = -\frac{1}{2}\sum_{x=n}^{m}\frac{[A_x - R_x q_{x+b}(\boldsymbol{\alpha})]^2}{r_x R_x q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\alpha)]} = -\frac{1}{2}\chi^2(\boldsymbol{\alpha}) \tag{6.2.23}$$

where $\chi^2(\boldsymbol{\alpha})$ is given by equation (6.2.8) above. This indicates why minimizing χ^2 produces a solution which is usually close to that obtained by maximizing the likelihood. Note that, by using either of the approximations $L_2(\boldsymbol{\alpha})$ and$L_3(\boldsymbol{\alpha})$, we may obtain maximum likelihood estimates, even in the situation when there are duplicates. These approximations are alternatives to the earlier empirical method of allowing for duplicates, whereby exposures and deaths are 'scaled down' (by division by the variance ratio at each age).

In § 6.5 below we indicate how we may obtain q-type rates when the available exposures are in 'central' form.

6.3 *Graduation of μ-type rates (using central exposures)*

Here we let

$$\mu_x = \mu_x(\boldsymbol{\alpha}) = f_{\boldsymbol{\alpha}}(x)$$

where, as before, $f_{\boldsymbol{\alpha}}(x)$ is a given function. Note that $R_x (n \leqslant x \leqslant m)$ now denotes the *central* exposure between exact age $x + b$ and exact age $x + b + 1$ and A_x is the corresponding number of actual deaths.

For practical purposes a Poisson distribution provides a suitable model for the distribution of A_x (see Sverdrup, 1965). In this situation we therefore assume that

$$A_x \sim P\left(R_x \int_{x+b}^{x+b+1} \mu_s(\boldsymbol{\alpha})\mathrm{d}s\right) \tag{6.3.1}$$

the notation $P(\lambda)$ being used to denote a Poisson distribution with parameter λ.

Letting

$$g_{\boldsymbol{\alpha}}(x) = \int_{x+b}^{x+b+1} f_{\boldsymbol{\alpha}}(s)\mathrm{d}s$$

we see that the Poisson parameter for the distribution of A_x is simply $R_x g_{\boldsymbol{\alpha}}(x)$, which (at least in principle) is a known function. In practice, however, exact integration of the function $f_{\boldsymbol{\alpha}}(x)$ may be difficult and approximate integration time-consuming. Accordingly, when estimating the vector $\boldsymbol{\alpha}$, it is more convenient to approximate the integral in this last expression by the value of the integrand at the mid-point of the range of integration. Thus for practical purposes we assume that the distribution for A_x is Poisson with parameter

$$\lambda_x = R_x \mu_{x+b+1/2}(\boldsymbol{\alpha}) \tag{6.3.2}$$

If exact integration of $f_{\boldsymbol{\alpha}}(x)$ is easily carried out, we may, of course, replace λ_x by

$$\lambda_x^* = R_x g_{\boldsymbol{\alpha}}(x) \tag{6.3.3}$$

For the remainder of this section, however, we shall assume that the Poisson parameter is given by equation (6.3.2). (It is a simple matter to adopt the alternative expression (6.3.3) when this can be readily calculated.)

(a) *Maximum likelihood*

When there are no duplicates the likelihood function is

$$L(\boldsymbol{\alpha}) = \prod_{x=n}^{m} \exp(-\lambda_x) \frac{(\lambda_x)^{A_x}}{A_x!} \tag{6.3.4}$$

Since

$$\log L(\boldsymbol{\alpha}) = \sum_{x=n}^{m} \{-\lambda_x + A_x \log \lambda_x - \log(A_x!)\}$$

and A_x and R_x do not depend on $\boldsymbol{\alpha}$, maximizing $L(\boldsymbol{\alpha})$ is equivalent to maximizing

$$L_1(\boldsymbol{\alpha}) = \sum_{x=n}^{m} \{-R_x \mu_{x+b+1/2}(\boldsymbol{\alpha}) + A_x \log \mu_{x+b+1/2}(\boldsymbol{\alpha})\} \tag{6.3.5}$$

In (c) below we describe one way of allowing for duplicates.

(b) *Minimum* χ^2

Since λ_x equals both the mean and variance of the distribution for A_x, the relative deviation at age x (when there are no duplicates) is

$$z_x = [A_x - \lambda_x]/(\lambda_x)^{1/2}$$

and, accordingly, we define

$$\chi^2(\boldsymbol{\alpha}) = \sum_{x=n}^{m} z_x^2 = \sum_{x=n}^{m} \frac{[A_x - R_x \mu_{x+b+1/2}(\boldsymbol{\alpha})]^2}{R_x \mu_{x+b+1/2}(\boldsymbol{\alpha})} \tag{6.3.6}$$

In the absence of duplicates $\boldsymbol{\alpha}$ is chosen to minimize this last expression.

When there are duplicates, A_x is the number of *policies* which become claims at age x during the observation period and R_x is the central exposure *based on policies*. In this case we may write

$$A_x = \sum_i i\theta_x^i \tag{6.3.7}$$

where θ_x^i is the number of *policyholders* who die at age x and have i policies. (Note that $\theta_x^1, \theta_x^2, \theta_x^3, \ldots$ relate to completely distinct groups of lives and are therefore the values of independent random variables.) Similarly we may write

$$R_x = \sum_i iT_x^i \tag{6.3.8}$$

where T_x^i is the central exposure *based on lives*, arising from those cases for which the policyholder has i policies.

Note now that the distribution for θ_x^i is Poisson with parameter

$$T_x^i \mu_{x+b+1/2}(\boldsymbol{\alpha}). \tag{6.3.9}$$

(This is exactly analogous to the expression (6.3.2) above.)

Clearly

$$E[A_x] = E\left[\sum_i i\theta_x^i\right] = \sum_i iE[\theta_x^i]$$

$$= \sum_i iT_x^i \mu_{x+b+1/2}(\boldsymbol{\alpha})$$

(Poisson distribution and equation (6.3.9) above)

$$= R_x \mu_{x+b+1/2}(\boldsymbol{\alpha}) \tag{6.3.10}$$

Thus, for an investigation based on policies, the existence of duplicates does not change the expected value of the number of claims. The variance, however, *is* altered. This is so, since

$$\text{Var}[A_x] = \text{Var}\left[\sum_i i\theta_x^i\right] = \sum_i i^2 \text{Var}[\theta_x^i]$$

(since $\theta_x^1, \theta_x^2, \ldots$ are independent)

$$= \sum_i i^2 T_x^i \mu_{x+b+1/2}(\boldsymbol{\alpha})$$

(Poisson distribution and equation (6.3.9) above)

$$= \frac{\sum_i i^2 T_x^i}{\sum_i iT_x^i} \cdot \sum_i iT_x^i \mu_{x+b+1/2}(\boldsymbol{\alpha})$$

$$= \frac{\sum_i i^2 T_x^i}{\sum_i iT_x^i} R_x \mu_{x+b+1/2}(\boldsymbol{\alpha})$$

(by equation (6.3.8) above)

$$= r_x R_x \mu_{x+b+1/2}(\boldsymbol{\alpha}) \tag{6.3.11}$$

where

$$r_x = \sum_i i^2 T_x^i \Big/ \sum_i iT_x^i \tag{6.3.12}$$

Provided that mortality and lapse rates do not depend on the number of policies held (and in most practical situations this may be a reasonable assumption), T_x^i will be proportional to f_x^i, the proportion of policyholders at age x who have i policies.

In this situation equation (6.3.12) then implies that

$$r_x = \sum_i i^2 f_x^i \Big/ \sum_i i f_x^i \qquad (6.3.13)$$

as before.

When allowance is made for duplicates $\boldsymbol{\alpha}$ is therefore chosen to minimize

$$\chi^2(\boldsymbol{\alpha}) = \frac{[A_x - R_x \mu_{x+b+1/2}(\boldsymbol{\alpha})]^2}{r_x R_x \mu_{x+b+1/2}(\boldsymbol{\alpha})} \qquad (6.3.14)$$

(c) *Maximum likelihood approximations and allowance for duplicates*

As for q-type rates, with the known mean and variance we may use the normal density function to approximate the likelihood. In this case the function to be maximized is

$$L(\boldsymbol{\alpha}) = \prod_{x=n}^{m} \frac{1}{[2\pi r_x \lambda_x]^{1/2}} \exp\left\{-\frac{(A_x - \lambda_x)^2}{2r_x\lambda_x}\right\} \qquad (6.3.15)$$

where λ_x and r_x are given by expressions (6.3.2) and (6.3.13) above.

Note that

$$\log L(\alpha) = -\frac{1}{2}\sum_{x=n}^{m}\left\{\log(2\pi r_x\lambda_x) + \frac{(A_x - \lambda_x)^2}{r_x\lambda_x}\right\}$$

Omitting the terms which do not depend on $\boldsymbol{\alpha}$, we see that the function to be maximized is

$$L_2(\boldsymbol{\alpha}) = -\frac{1}{2}\sum_{x=n}^{m}\left\{\log \mu_{x+b+1/2}(\boldsymbol{\alpha}) + \frac{[A_x - R_x\mu_{x+b+1/2}(\boldsymbol{\alpha})]^2}{r_x R_x \mu_{x+b+1/2}(\boldsymbol{\alpha})}\right\} \qquad (6.3.16)$$

Empirical experiments indicate that the value of the logarithmic term on the right-hand side of this last equation is much less sensitive to changes in $\boldsymbol{\alpha}$ than the other term—at least near the optimal solution. This means that a good approximation to the optimal solution can usually be found by maximizing

$$L_3(\boldsymbol{\alpha}) = -\frac{1}{2}\sum_{x=n}^{m}\frac{[A_x - R_x\mu_{x+b+1/2}(\boldsymbol{\alpha})]^2}{r_x R_x\mu_{x+b+1/2}(\boldsymbol{\alpha})} = -\frac{1}{2}\chi^2(\boldsymbol{\alpha}) \qquad (6.3.17)$$

where $\chi^2(\boldsymbol{\alpha})$ is defined by equation (6.3.14) above.

As in §6.2 above (for q-type rates), these approximations allow maximum likelihood methods to be used in a graduation of μ-type rates when there are duplicates. As before, this last equation indicates why minimizing $\chi^2(\boldsymbol{\alpha})$ is likely to lead to a similar graduation to that obtained by the method of maximum likelihood.

6.4 *Graduation of* m-*type rates (using central exposures)*

Here we let

$$m_x = m_x(\boldsymbol{\alpha}) = f_{\boldsymbol{\alpha}}(x)$$

where $f_{\boldsymbol{\alpha}}(x)$ is the given function to be used for the graduation.

The situation is essentially identical to that of the previous section, R_x again being the central exposure at age x. The Poisson model remains appropriate.

As before, if there are no duplicates, the Poisson parameter for the distribution of A_x is

$$R_x \int_{x+b}^{x+b+1} \mu_s \mathrm{d}s.$$

In a graduation of m-type rates we assume that

$$\int_{x-b}^{x-b+1} \mu_s \mathrm{d}s = m_{x+b} \tag{6.4.1}$$

For practical purposes this is a reasonable approximation, which differs only slightly from the correct expression for m_{x+b}, namely

$$\int_{x-b}^{x-b-1} l_s \mu_s \mathrm{d}s \Big/ \int_{x+b}^{x+b+1} l_s \mathrm{d}s.$$

The approximation (6.4.1) having been made, we may now repeat the discussion of the previous section—simply replacing $\mu_{x+b+1/2}(\boldsymbol{\alpha})$ by $m_{x+b}(\boldsymbol{\alpha})$ throughout. For completeness it is convenient to record the results corresponding to those of § 6.3.

(a) *Maximum likelihood*

In the absence of duplicates the function to be maximized is

$$L_1(\boldsymbol{\alpha}) = \sum_{x=n}^{m} \{- R_x m_{x+b}(\boldsymbol{\alpha}) + A_x \log m_{x+b}(\boldsymbol{\alpha})\} \tag{6.4.2}$$

(b) *Minimum χ^2*

Let r_x be defined by equation (6.3.13) above. The function to be minimized is

$$\chi^2(\boldsymbol{\alpha}) = \sum_{x=n}^{m} \frac{[A_x - R_x m_{x+b}(\boldsymbol{\alpha})]^2}{r_x R_x m_{x+b}(\boldsymbol{\alpha})} \tag{6.4.3}$$

(c) *Maximum likelihood approximations and allowance for duplicates*

Using the normal approximation and allowing for duplicates, we may maximize either

$$L_2(\boldsymbol{\alpha}) \quad = \quad -\frac{1}{2}\sum_{x=n}^{m}\left\{\log m_{x+b}(\boldsymbol{\alpha}) + \frac{[A_x - R_x m_{x+b}(\boldsymbol{\alpha})]^2}{r_x R_x m_{x+b}(\boldsymbol{\alpha})}\right\} \qquad (6.4.4)$$

or

$$L_3(\boldsymbol{\alpha}) \quad = \quad -\frac{1}{2}\sum_{x=n}^{m}\frac{[A_x - R_x m_{x+b}(\boldsymbol{\alpha})]^2}{r_x \mathrm{R}_x m_{x+b}(\boldsymbol{\alpha})} \quad = \quad -\frac{1}{2}\chi^2(\boldsymbol{\alpha}). \qquad (6.4.5)$$

Remark In combination the approximations used in §§ 6.3 and 6.4 imply that

$$m_{x+b} \quad = \quad \mu_{x+b+1/2}$$

For almost all practical purposes this is a reasonable approximation. Its use means that formula graduations for μ-type or m-type rates are equivalent procedures. Each arises naturally from the other simply by an age adjustment of 1/2.

6.5 *Graduation of* q-*type rates (using central exposures)*

Although it is perhaps more natural with central exposures to consider a formula graduation for μ_x or m_x, as we have remarked in § 6.2 above it is perfectly feasible to proceed directly to a graduation of q-type rates in this case. We indicate below how this may be done, using the Poisson model of §§ 6.3 and 6.4 above.

Suppose then that we have available *central* exposures and that

$$q_x \quad = \quad q_x(\boldsymbol{\alpha}) \quad = \quad f_{\boldsymbol{\alpha}}(x)$$

where $f_{\boldsymbol{\alpha}}(x)$ is the given function to be used in the graduation.

Note that

$$\int_{x+b}^{x+b+1} \mu_s \mathrm{d}s \quad = \quad -\log(1 - q_{x+b}) \qquad (6.5.1)$$

so the parameter for the Poisson distribution of A_x (see equation (6.3.1) above) is

$$\lambda_x \quad = \quad -R_x \log(1 - q_{x+b}(\boldsymbol{\alpha}))$$

R_x being the central exposure at age x.

We may therefore repeat our previous discussion, simply replacing $\mu_{x+b+1/2}(\boldsymbol{\alpha})$ by $-\log(1 - q_{x+b}(\boldsymbol{\alpha}))$ throughout. Again for completeness, it is convenient, to record the results corresponding to those of § 6.3.

(a) *Maximum likelihood*

In the absence of duplicates the function to be maximized is

$$L_1(\boldsymbol{\alpha}) \quad = \quad \sum_{x=n}^{m}\{R_x \log(1 - q_{x+b}(\boldsymbol{\alpha})) + A_x \log[-\log(1 - q_{x+b}(\boldsymbol{\alpha}))]\} \qquad (6.5.2)$$

(b) *Minimum* χ^2

Let r_x be defined by equation (6.3.13) above.

The function to be minimized is

$$\chi^2(\boldsymbol{\alpha}) = -\sum_{x=n}^{m} \frac{[A_x + R_x \log(1 - q_{x+b}(\boldsymbol{\alpha}))]^2}{r_x R_x \log(1 - q_{x+b}(\boldsymbol{\alpha}))} \qquad (6.5.3)$$

(c) *Maximum likelihood approximations and allowance for duplicates*

Using the normal approximation and allowing for duplicates, we may maximize either

$$L_2(\boldsymbol{\alpha}) = -\frac{1}{2}\sum_{x=n}^{m}\left\{\log[-\log(1 - q_{x+b}(\boldsymbol{\alpha}))] - \frac{[A_x + R_x \log(1 - q_{x+b}(\boldsymbol{\alpha}))]^2}{r_x R_x \log(1 - q_{x+b}(\boldsymbol{\alpha}))}\right\} \qquad (6.5.4)$$

or

$$L_3(\boldsymbol{\alpha}) = -\frac{1}{2}\chi^2(\boldsymbol{\alpha}) \qquad (6.5.5)$$

where $\chi^2(\boldsymbol{\alpha})$ is given by equation (6.5.3) above.

7. METHODS OF OPTIMISATION

In Section 6 we have described in some detail the models underlying a graduation by formula in a variety of situations, either by maximum likelihood or by minimum χ^2. In each case we require to locate the point, $\hat{\boldsymbol{\alpha}}$, at which a function of several variables (i.e. minus the log-likelihood or χ^2) takes its minimum value. In certain situations the problem may involve *constraints*. For example, if the graduating function $f_{\boldsymbol{\alpha}}(x)$ is being fitted to q-type rates, then the optimal point $\hat{\boldsymbol{\alpha}}$ must be such that $0 < f_{\hat{\boldsymbol{\alpha}}}(x) < 1$ for all x in the relevant age-range. If the graduation is by splines, there is the constraint that the knot-positions must lie within the appropriate interval.

Except in cases too simple to be of use in practice, without some form of computer to assist our calculations we are unlikely to solve this problem. However, given even only limited computing power, we are generally able to make substantial progress—usually by locating one or more points at which the relevant function has a *local* minimum. The empirical procedure then adopted is to find as many local minima as possible, starting from widely different initial points, and to pick the lowest of these (subject to any constraints) as our solution $\hat{\boldsymbol{\alpha}}$. In the majority of applications this is an acceptable procedure, but one cannot eliminate completely the possibility that there may exist a 'lower' point. In a few limited situations (depending, for example, on the functional form being used as the graduation formula) further analysis may confirm our point as a global minimum, but in most cases such confirmation will not be possible. Moreover, at present little is known about finding global minima in general.

The most suitable computer method for any particular computational procedure may depend on a balance between various desirable and possibly conflicting objectives—speed of running, memory requirements, ready availability, and ease of use, amongst others. If computer library facilities are available, appropriate routines will be readily to hand. Otherwise one must write one's own routines, or use routines prepared, tested, and published by others. Fortunately a variety of such routines exists. For example, in Press *et al.* (1986) the reader will find several routines in both the FORTRAN and PASCAL languages.

It is inappropriate to discuss in detail here computer routines for the minimization of a function of several variables. It is, however, perhaps worth pointing out that such routines fall into two distinct categories—those which require the calculation of first partial derivatives and those which do not. Routines which require derivatives are somewhat more powerful than those requiring only function values. On the other hand the increased power of such routines may not always compensate for the additional calculation of derivatives. Methods which do not require the calculation of derivatives, such as the *downhill simplex method* of Nelder & Mead (1965), can be used in a wider variety of circumstances and may give equally satisfactory results. For maximum likelihood methods the relevant partial derivatives (if required) are easily obtained from the calculations in Appendix 1.

The interested reader may consult Conte & de Boor (1980) and Press *et al.* (1986) for further details.

8. THE INFORMATION MATRIX AND THE VARIANCE-COVARIANCE MATRIX FOR THE PARAMETERS TO BE ESTIMATED

8.1 Suppose now that our curve-fitting exercise is being done by a maximum likelihood method. The properties of maximum likelihood estimators are described in detail in most statistics textbooks. (The interested reader may refer, for example, to Kendall & Stuart (1979) for a discussion of the topic.) In this section, therefore, we give only an outline of some of the basic ideas, with particular reference to the situation described in § 6 above.

Suppose then that we observe a population for which the rate of mortality at exact age x (i.e. q_x, μ_x, or m_x depending upon the available data and the model we are using) is given by a function $f_{\boldsymbol{\alpha}}(x)$. Suppose further that the functional *form* of $f_{\boldsymbol{\alpha}}(x)$ is known, but that the values of the coefficients in the formula are not. The vector $\boldsymbol{\alpha}$ (which we sometimes call the 'true' vector to distinguish it from estimates) is thus unknown.

Let $L^*(\boldsymbol{\alpha})$ be the appropriate log-likelihood function as defined in § 6. If there are no duplicates, then normally we will let

$$L^*(\alpha) \;=\; L_1(\boldsymbol{\alpha}) \tag{8.1.1}$$

where $L_1(\boldsymbol{\alpha})$ is given by equation (6.2.4), (6.3.5), (6.4.2), or (6.5.2) as appropriate. If there are duplicates or if for some other reason it is more convenient to adopt one of the approximations for the likelihood, we will let

$$L^*(\boldsymbol{\alpha}) = L_2(\boldsymbol{\alpha}) \tag{8.1.2}$$

or

$$L^*(\boldsymbol{\alpha}) = L_3(\boldsymbol{\alpha}) \tag{8.1.3}$$

where $L_2(\boldsymbol{\alpha})$ and $L_3(\boldsymbol{\alpha})$ are the appropriate alternative approximations described in §§6.2–6.5 above.

Let $\hat{\boldsymbol{\alpha}}$ be the vector which maximizes $L^*(\boldsymbol{\alpha})$. Then $\hat{\boldsymbol{\alpha}}$ is taken as our 'best' estimate for the coefficients in the formula for the death rate. Since the definition of $L^*(\boldsymbol{\alpha})$ involves the data set $\{A_x\}$, the value of $\hat{\boldsymbol{\alpha}}$ depends on the numbers of deaths which occur at each age. Thus $\hat{\boldsymbol{\alpha}}$ should be considered as one particular value from a random distribution—i.e. a different set of values $\{A_x\}$ would lead to a different value of $\hat{\boldsymbol{\alpha}}$.

The important point about the estimator $\hat{\boldsymbol{\alpha}}$ is that asymptotically it is unbiased and of minimum variance. In particular, as the exposures increase, the expected value of $\hat{\boldsymbol{\alpha}}$ tends to $\boldsymbol{\alpha}$ (the true but unknown coefficient vector). We may therefore write

$$E[\hat{\boldsymbol{\alpha}}] \rightarrow \boldsymbol{\alpha} \tag{8.1.4}$$

the limit being considered as the smallest exposure to risk tends to infinity. One cannot specify precisely the speed of convergence to the limit. Empirical experiments, however, show that in most large scale investigations the speed of convergence (combined with the minimum variance property) is such that $\hat{\boldsymbol{\alpha}}$ does indeed provide an acceptable estimate for $\boldsymbol{\alpha}$.

8.2 In addition to estimating the coefficient vector $\boldsymbol{\alpha}$ we will find it extremely valuable to know (at least to a reasonable degree of accuracy) the corresponding variance-covariance matrix for the estimator $\hat{\boldsymbol{\alpha}}$. This matrix will, for example, enable us to compare different experiences (for which we have graduations) and to ascertain whether or not there is a significant difference between them. (See §11 below.) One of the attractive features of maximum-likelihood estimators is that, at least asymptotically, they enable us to determine the corresponding variance-covariance matrix. The starting point for our calculations is the matrix $H^*(\boldsymbol{\alpha})$. This has $(i, j)^{\text{th}}$ entry, $H^*_{ij}(\boldsymbol{\alpha})$, equal to *minus* the second partial derivative of $L^*(\boldsymbol{\alpha})$ with respect to α_i and α_j, evaluated at the (unknown) point $\boldsymbol{\alpha}$. Thus

$$H^*_{ij}(\boldsymbol{\alpha}) = \frac{-\,\partial^2 L^*(\boldsymbol{\alpha})}{\partial\alpha_i \partial\alpha_j} \tag{8.2.1}$$

Note that the entries of $H^*(\boldsymbol{\alpha})$ are random variables, which in general will depend on the outcome actually observed—in our case on the numbers of deaths at each age. In Appendix 1 we derive the value of $H^*_{ij}(\boldsymbol{\alpha})$ for each of the models described above.

The *information matrix* $H(\boldsymbol{\alpha})$ is defined to be the expected value of $H^*(\boldsymbol{\alpha})$. Thus

$$H(\boldsymbol{\alpha}) = E[H^*(\boldsymbol{\alpha})] \tag{8.2.2}$$

Standard properties of maximum likelihood estimators show that asymptotically

$$\text{Cov}(\hat{\boldsymbol{\alpha}}) = [H(\boldsymbol{\alpha})]^{-1} \tag{8.2.3}$$

where $\text{Cov}(\hat{\boldsymbol{\alpha}})$ denotes the variance-covariance matrix for $\hat{\boldsymbol{\alpha}}$.

By combining these last equations we immediately obtain the well-known asymptotic result

$$\text{Cov}(\hat{\boldsymbol{\alpha}}) = \left[-E\left[\frac{\partial^2 L^*(\boldsymbol{\alpha})}{\partial\alpha_i\partial\alpha_j}\right]\right]^{-1} \tag{8.2.4}$$

One difficulty with this last equation is that it requires the matrix of second partial derivatives to be evaluated at the *unknown* point $\boldsymbol{\alpha}$. In practice this is not a serious problem, as $\hat{\boldsymbol{\alpha}}$ generally provides a good estimate for $\boldsymbol{\alpha}$ and, accordingly, by evaluating the right-hand side of equation (8.2.4) at $\hat{\boldsymbol{\alpha}}$ we generally obtain an acceptable approximation to $\text{Cov}(\hat{\boldsymbol{\alpha}})$. In this case the variance-covariance matrix is estimated simply as

$$\text{Cov}(\hat{\boldsymbol{\alpha}}) = [H(\hat{\boldsymbol{\alpha}})]^{-1} \tag{8.2.5}$$

Alternatively, one may not take expected values and simply work with *actual* values (at the point $\hat{\boldsymbol{\alpha}}$). This gives as estimate for the variance-covariance matrix

$$\text{Cov}(\hat{\boldsymbol{\alpha}}) = [H^*(\hat{\boldsymbol{\alpha}})]^{-1} \tag{8.2.6}$$

In general these two estimates of $\text{Cov}(\hat{\boldsymbol{\alpha}})$ are distinct. For most practical purposes, however, either estimate may be used.

One advantage of using the expected value is that in the asymptotic limit in certain cases the second order partial derivatives vanish, leaving only products of first derivatives. For these cases the entries of $\text{Cov}(\hat{\boldsymbol{\alpha}})$ can be found simply in terms of the *first* partial derivatives of $L^*(\boldsymbol{\alpha})$. This has advantages from the computational viewpoint. This point is discussed further in Appendix 1, where we give detailed calculations for the information matrix in all cases.

In § 8.3 we summarise the results from Appendix 1.

8.3 For completeness we give below the information matrix (corresponding to the use of $L_1(\boldsymbol{\alpha})$, $L_2(\boldsymbol{\alpha})$, or $L_3(\boldsymbol{\alpha})$ for the log-likelihood) for each of the models (q, μ, or m) previously discussed. For $k = 1, 2, 3$ let $H^k(\boldsymbol{\alpha})$ be the information matrix when $L^*(\boldsymbol{\alpha}) = L_k(\boldsymbol{\alpha})$. Thus $H^k(\boldsymbol{\alpha})$ has $(i, j)^{\text{th}}$ entry

$$H_{ij}^k(\boldsymbol{\alpha}) = -E\left[\frac{\partial^2 L_k(\boldsymbol{\alpha})}{\partial\alpha_i\partial\alpha_j}\right] \tag{8.3.1}$$

The corresponding estimate for $\text{Cov}(\hat{\boldsymbol{\alpha}})$ (see equation (8.2.5) above) is

$$\mathrm{Cov}_k(\hat{\boldsymbol{\alpha}}) = [H^k(\hat{\boldsymbol{\alpha}})]^{-1} \tag{8.3.2}$$

For notational simplification, when giving the values of the information matrix for each of the models described in § 6, we write R, r, q, μ, and m to denote R_x, r_x, $q_{x+b}(\boldsymbol{\alpha})$, $\mu_{x+b+1/2}(\boldsymbol{\alpha})$, and $m_{x+b}(\boldsymbol{\alpha})$ respectively.

For each model, in the definition of the information matrix given below the summation is over the range of ages appropriate to the definition of $L^*(\boldsymbol{\alpha})$.

(a) *q-type rates (initial exposures - see § 6.2)*

(i) $$H^1_{ij}(\boldsymbol{\alpha}) = \sum_x \frac{R}{q(1-q)} \frac{\partial q}{\partial \alpha_i} \frac{\partial q}{\partial \alpha_j} \tag{8.3.3}$$

(Appendix 1 equation (8).)

(ii) $$H^2_{ij}(\boldsymbol{\alpha}) = \sum_x \left\{ \frac{(1-2q)^2}{2q^2(1-q)^2} + \frac{R}{rq(1-q)} \right\} \frac{\partial q}{\partial \alpha_i} \frac{\partial q}{\partial \alpha_j} \tag{8.3.4}$$

(Appendix 1 equation (16).)

(iii) $$H^3_{ij}(\alpha) = \sum_x \left\{ -\frac{1-2q}{2q(1-q)} \frac{\partial^2 q}{\partial \alpha_i \partial \alpha_j} + \left[\frac{1-3q+3q^2}{q^2(1-q)^2} + \frac{R}{rq(1-q)} \right] \frac{\partial q}{\partial \alpha_i} \frac{\partial q}{\partial \alpha_j} \right\} \tag{8.3.5}$$

(Appendix 1 equation (21).)

(b) *μ-type rates (central exposures - see § 6.3)*

(i) $$H^1_{ij}(\boldsymbol{\alpha}) = \sum_x \frac{R}{\mu} \frac{\partial \mu}{\partial \alpha_i} \frac{\partial \mu_j}{\partial \alpha_j} \tag{8.3.6}$$

(Appendix 1 equation (27).)

(ii) $$H^2_{ij}(\boldsymbol{\alpha}) = \sum_x \left\{ \frac{1}{2\mu^2} + \frac{R}{r\mu} \right\} \frac{\partial \mu}{\partial \alpha_i} \frac{\partial \mu}{\partial \alpha_j} \tag{8.3.7}$$

(Appendix 1 equation (32).)

(iii) $$H^3_{ij}(\boldsymbol{\alpha}) = \sum_x \left\{ -\frac{1}{2\mu} \frac{\partial^2 \mu}{\partial \alpha_i \partial \alpha_j} + \left[\frac{1}{\mu^2} + \frac{R}{r\mu} \right] \frac{\partial \mu}{\partial \alpha_i} \frac{\partial \mu}{\partial \alpha_j} \right\} \tag{8.3.8}$$

(Appendix 1 equation (37).)

(c) *m-type rates (central exposures - see § 6.4)*

(i) $$H^1_{ij}(\boldsymbol{\alpha}) = \sum_x \frac{R}{m} \frac{\partial m}{\partial \alpha_i} \frac{\partial m}{\partial \alpha_j} \tag{8.3.9}$$

(Appendix 1 equation (38).)

(ii) $$H_{ij}^2(\boldsymbol{\alpha}) = \sum_x \left\{\frac{1}{2m^2} + \frac{R}{rm}\right\} \frac{\partial m}{\partial \alpha_i}\frac{\partial m}{\partial \alpha_j} \quad (8.3.10)$$

(Appendix 1 equation (39).)

(iii) $$H_{ij}^3(\boldsymbol{\alpha}) = \sum_x \left\{- \frac{1}{2m}\frac{\partial^2 m}{\partial \alpha_i \partial \alpha_j} + \left[\frac{1}{m^2} + \frac{R}{rm}\right] \frac{\partial m}{\partial \alpha_i}\frac{\partial m}{\partial \alpha_j}\right\} \quad (8.3.11)$$

(Appendix 1 equation (40).)

(d) *q-type rates (central exposures - see § 6.5)*

(i) $$H_{ij}^1(\boldsymbol{\alpha}) = \sum_x \left\{- \frac{R}{(1-q)^2[\log(1-q)]}\frac{\partial q}{\partial \alpha_i}\frac{\partial q}{\partial \alpha_j}\right\} \quad (8.3.12)$$

(Appendix 1 equation (47).)

(ii) $$H_{ij}^2(\boldsymbol{\alpha}) = \sum_x \frac{1}{(1-q)^2}\left[\frac{1}{2[\log(1-q)]^2} - \frac{R}{r\log(1-q)}\right]\frac{\partial q}{\partial \alpha_i}\frac{\partial q}{\partial \alpha_j} \quad (8.3.13)$$

(Appendix 1 equation (49).)

(iii) $$H_{ij}^3(\boldsymbol{\alpha}) = \sum_x \left\{\frac{1}{2(1-q)\log(1-q)}\frac{\partial^2 q}{\partial \alpha_i \partial \alpha_j} + \frac{1}{(1-q)^2}\left[\frac{1}{2\log(1-q)} + \frac{1}{[\log(1-q)]^2} - \frac{R}{r\log(1-q)}\right]\frac{\partial q}{\partial \alpha_i}\frac{\partial q}{\partial \alpha_j}\right\} \quad (8.3.14)$$

(Appendix 1 equation (51).)

9. TESTS OF A GRADUATION

9.1 *Deviations and relative deviations*

Once we have chosen a particular functional form, $f_{\boldsymbol{\alpha}}(x)$ to represent q_x, μ_x or m_x, as appropriate, and found those values of the parameters (i.e. the vector $\hat{\boldsymbol{\alpha}}$) that maximise the likelihood or minimise χ^2, as desired, we wish to test the resulting graduation to see whether we can accept it as a reasonable representation of the mortality experience of the investigation. At this point we define, as usual, for each integer observed age, x, the expected number of deaths, E_x, and the corresponding variance of the number of deaths, V_x, as follows. Where there are duplicates we define the appropriate variance ratio as r_x, and if there are no duplicates $r_x = 1$. Let b be as defined in § 3.1.

If we have graduated q_x (with initial exposures),

$$E_x = R^i_x q_{x+b} \tag{9.1.1}$$

and

$$V_x = r_x R^i_x q_{x+b} (1 - q_{x+b}); \tag{9.1.2}$$

if we have graduated μ_x,

$$E_x = R^c_x \mu_{x+b+1/2} \tag{9.1.3}$$

and

$$V_x = r_x R^c_x \mu_{x+b+1/2}; \tag{9.1.4}$$

if we have graduated m_x,

$$E_x = R^c_x m_{x+b} \tag{9.1.5}$$

and

$$V_x = r_x R^c_x m_{x+b}; \tag{9.1.6}$$

if we have graduated q_x (with central exposures),

$$E_x = - R^c_x \log (1 - q_{x+b}) \tag{9.1.7}$$

and

$$V_x = - r_x R^c_x \log (1 - q_{x+b}) \tag{9.1.8}$$

We then define the deviation at each integral age as

$$\text{Dev}_x = A_x - E_x \tag{9.1.9}$$

and the relative deviation as

$$z_x = \frac{\text{Dev}_x}{\{V_x\}^{1/2}} \tag{9.1.10}$$

As stated in CMI Committee (1974) the questions to be asked are whether the deviations or the relative deviations are

(a) randomly distributed (when their sequence is considered),

and

(a) distributed in accordance with the assumptions inherent in the model used, i.e. normal, Poisson, or binomial.

Some of the tests that can be used are familiar, and we shall not dwell on these. Others have not been used before in the actuarial literature. Some of the tests are applicable also to test whether an observed experience is compatible with some other set of rates, whether based on a mathematical formula or not, and are therefore appropriate for testing, for example, whether the experience at one select duration is satisfactorily represented by the graduated rates at some other duration. The comparison of two experiences is discussed further in § 13.

9.2 *Grouping of ages*

Most of the tests rely on the assumption that the numbers of deaths at each age are approximately normally distributed. This is a reasonable assumption only if the expected numbers of deaths at each age are sufficiently large, such as at least 5 or so. In almost all investigations the numbers at each age may be less than this at the extreme ages under consideration, and in some investigations the data may be scanty at more than just the extreme ages. It is therefore useful to group neighbouring ages so that the sum of the expected deaths in the group is at least some number, k, such as 5. This can be done by an algorithm such as the following: starting at the lowest age in the table of data, consecutive ages are grouped until the sum of the expected deaths in the group is not less than k; a group is completed at that point; the next group is then started; a group may consist of a single age; at the end of the table, if the last (incomplete) group has fewer than k expected deaths, it is added to the last completed group. Different algorithms for grouping may produce different groups; one could start at the highest age and run down the table, or at both ends and run towards the central region (provided that the centre of the table is sufficiently dense), or at the centre and run out. It probably makes little difference which algorithm is adopted, but it is always desirable to choose the algorithm for grouping before examining the results, so that there is no possibility that the method of grouping is chosen to produce favourable or unfavourable results. In each of the examples discussed in § 15 below the grouping method has been as described above, with $k = 5$ and with grouping from the youngest age in the table upwards; any residual group at the highest age is added back to the group formed by the next lower ages.

The justification for grouping is readily explained. If the number of deaths at age x is assumed to be distributed (whether normally or not) with mean E_x and variance V_x, then the number of deaths at ages x_1 to x_2 is distributed with mean

$$E_{x_1,x_2} = \sum_{r=x_1}^{x_2} E_r \tag{9.2.1}$$

and variance

$$V_{x_1,x_2} = \sum_{r=x_1}^{x_2} V_r \tag{9.2.2}$$

As E_{x_1,x_2} increases, the distribution of the number of deaths in the group from x_1 to x_2 tends to the normal distribution. The actual number of deaths in the group is denoted by

$$A_{x_1,x_2} = \sum_{r=x_1}^{x_2} A_r \tag{9.2.3}$$

the deviation by

$$\text{Dev}_{x_1,x_2} = A_{x_1,x_2} - E_{x_1,x_2} \tag{9.2.4}$$

and the relative deviation by

$$z_{x_1,x_2} = \frac{\text{Dev}_{x_1,x_2}}{\{V_{x_1,x_2}\}^{1/2}} \tag{9.2.5}$$

Further discussion about tests will generally assume that the data has been grouped before the test is applied, and it will be assumed that there are N distinct ages or age groups.

9.3 *Signs test*

The first test to use is the signs test. This test considers only the signs (positive or negative) of the deviations. If the observed numbers of deaths come from the experience implied by the graduated rates, and if at each age the distribution of the number of deaths has its mean equal to its median (as is the case with the normal distribution, but not the Poisson distribution), then for each deviation the probabilities that its sign is positive or negative are both equal to one half. The number of positive (or negative) signs is therefore binomially distributed as $B(N, \frac{1}{2})$. If the number of positive signs is denoted by NP, then

$$P(NP = r) = \frac{N!}{r!(N - r)!} \cdot \frac{1}{2^N} \tag{9.3.1}$$

The expected number of positive signs is $N/2$ and the variance is $N/4$. If N is large, then it can be assumed that the number of positive signs is approximately normally distributed, but it is not difficult, when the graduation tests are being carried out by computer, to calculate from the binomial distribution the exact probability that the number of positive signs would not exceed the observed number. We denote this probability by p(pos).

If the value of p(pos) is too low, such as less than ·025, then the observed number of positive deviations is unexpectedly small, and if it is too high, such as greater than ·975, then the observed number of negative signs is unexpectedly small. In either case, the graduated rates are too far to one side or the other of the observed rates, and are unlikely to be a satisfactory representation of the experience. In practice we have found that, where the parameters of the graduated rates have been fitted by one of the optimising methods described above, it is rare for this test to show a significantly low or high probability, even if the graduation is obviously unsatisfactory in other respects. For further discussion of the signs test and the more sensitive Wilcoxon signed ranks test see Larson (1982) Chapter 10.

9.4 *Runs test*

A second non-parametric test is the runs test or the sign-change test. If the numbers of deaths at each age were distributed according to the normal model, the deviations at successive ages would be independent, and the signs of the deviations would be randomly distributed, with neither too many nor too few runs of successive deviations with the same sign. If the number of positive signs is n_1 and the number of negative signs is n_2 (with $n_1 + n_2 = N$), and if the signs are arranged at random within the sequence of groups, then the distribution of NR, the number of runs of one or more consecutive deviations with the same sign, is known. The probabilities of different values of NR are given by

$$P(NR = r) = \frac{2(n_1 - 1)!}{(k-1)!\,(n_1 - k)!} \frac{(n_2 - 1)!}{(k-1)!\,(n_2 - k)!} \frac{n_1!\,n_2!}{N!} \tag{9.4.1}$$

if $r = 2k$ (i.e. r is even), and

$$P(NR = r) = \frac{(n_1 - 1)!}{(k-1)!\,(n_1 - k)!} \frac{(n_2 - 1)!}{k!\,(n_2 - 1 - k)!} \frac{n_1!\,n_2!}{N!} + \frac{(n_1 - 1)!}{k!\,(n_1 - k - 1)!} \frac{(n_2 - 1)!}{(k-1)!\,(n_2 - k)!} \frac{n_1!\,n_2!}{N!} \tag{9.4.2}$$

if $r = 2k + 1$ (i.e. r is odd)

The mean number of runs is

$$E[NR] = \frac{2n_1 n_2}{N} + 1 \tag{9.4.3}$$

and the variance is

$$\text{Var}[NR] = \frac{2n_1 n_2 (2n_1 n_2 - N)}{N^2 (N - 1)} \tag{9.4.4}$$

If both n_1 and n_2 are larger than about 20 then the number of runs is approximately normally distributed, but it is not difficult with a computer to calculate the exact distribution of NR, and hence the probability, p(runs), that the value of NR is less than or equal to the observed value. A low value of p(runs) means too few runs, which is typical of a graduation that is too straight compared with the observed values, cutting across waves or bends in the observed rates. In this case a formula with a larger number of parameters may be needed to give a satisfactory fit. This can be achieved, for example, by increasing the order of the polynomials in a GM(r, s) formula. However, some otherwise satisfactory graduations show rather low values for p(runs) and examination of the observed rates shows that a formula with a very much larger number of parameters would be needed if the number of runs were to be increased sufficiently. A too high value for p(runs) rarely occurs, but when it does it indicates that the graduation follows the observed experience too slavishly, weaving on either side of the observed rates, and is an indication of over-fitting. A satisfactory value for p(runs) shows that the graduated rates run comfortably down the middle of the observed rates. See Larson (1982) Chapter 10 or Fisz (1963) Chapter 11 for further discussion of this test.

9.5 *Kolmogorov–Smirnov test*

A further non-parametric test that can be applied is the Kolmogorov–Smirnov test. This test considers the distribution of the maximum absolute deviation between two cumulative distributions. It has been described by the CMI Committee (1986), where it is used to compare the cumulative distributions of the exposed to risk or actual deaths in two successive years. In graduation we use it to compare the distributions of actual and expected deaths.

Consider the distributions of the actual deaths and the expected deaths by age from x_1 to x_2, viz $\{A_x\}$ and $\{E_x\}$. Define the total actual deaths as

$$A = \sum_{r=x_1}^{x_2} A_r \tag{9.5.1}$$

and the total expected deaths as

$$E = \sum_{r=x_1}^{x_2} E_r \tag{9.5.2}$$

Define the cumulative distributions

$$F(x) = \frac{\sum_{r=x_1}^{x} A_r}{A} \tag{9.5.3}$$

and

$$G(x) = \frac{\sum_{r=x_1}^{x} E_r}{E} \tag{9.5.4}$$

The maximum absolute difference between the cumulative distributions is

$$D = \operatorname*{Max}_{x} |F(x) - G(x)| \tag{9.5.5}$$

The Kolmogorov–Smirnov statistic

$$KS = D\left\{\frac{AE}{A + E}\right\}^{1/2} \tag{9.5.6}$$

has a known distribution, so the probability, $p(KS)$ say, of a value as large as or larger than that actually obtained can be calculated or derived from tables of the distribution. See Fisz (1963) or Durbin (1973) for further discussion of this test.

If the expected numbers of deaths are based on a graduation produced by one of the optimising methods described above then E is very often equal to A. (See Appendix 2.) This result strictly depends on the formula used to represent the mortality rates. To show that the Kolmogorov–Smirnov statistic is related to the maximum absolute value of the cumulative sum of the deviations we put

$$C(x) = \sum_{r=x_1}^{x} \text{Dev}_r \tag{9.5.7}$$

$$= \sum_{r=x_1}^{x} (A_r - E_r) \quad (9.5.8)$$

$$= A \cdot F(x) - E \cdot G(x) \quad (9.5.9)$$

Thus, if $E = A$,

$$C(x) = A[F(x) - G(x)] \quad (9.5.10)$$

$$\text{Max } |C(x)| = \text{Max } |F(x) - G(x)| \, A \quad (9.5.11)$$

$$= DA \quad (9.5.12)$$

and

$$KS = D \left\{ \frac{A^2}{2A} \right\}^{1/2}$$

$$= D \left\{ \frac{A}{2} \right\}^{1/2} \quad (9.5.13)$$

$$= \frac{\text{Max}|C(x)|}{\{2A\}^{1/2}} \quad (9.5.14)$$

Thus a test traditionally used by actuaries, as described for example by Benjamin & Pollard (1970) p. 231, is seen to be related to a more general result, well known by statisticians.

Unfortunately the Kolmogorov–Smirnov test used in this way is fully valid only when the rates on which the expected deaths are based are independent of the actual deaths, which is clearly not the case when the expected deaths are based on graduated rates derived from the actual deaths. In this case the cumulative expected deaths should be much closer to the cumulative actual deaths than the test implies. Even so, the test remains a useful negative one. Too high a value for the maximum deviation (or equivalently too low a value for p(KS)) indicates that the graduation is certainly not a satisfactory one. In practice we have found that satisfactory graduations often produce values of p(KS) higher than ·9, sometimes exceeding ·99.

9.6 *Serial correlation test*

A test that has been used previously by the CMI Committee (1974) is the serial correlation test. Each relative deviation, z_x, has approximately a unit normal distribution, that is, it is normally distributed with mean zero and variance 1. The values of z_x form a sequence, and the serial correlation coefficients, $r_j, j = 1, 2, 3, \ldots$, (that is the correlation coefficients between the values of z_x and z_{x-j}) can be calculated. If z_x and z_{x-j} are independent, then r_j is normally distributed with zero mean and variance $1/N$, where N is the number of ages or age groups. One can therefore calculate $t_j = r_j/N^{-1/2}$ and compare this with a unit normal distribution. Too high a value indicates that successive values of z_x are too closely related, and the graduation is less than satisfactory. A low value for p(runs), indicating that the number of runs is too few, is often associated with high values for t_j for low values of j (say $j = 1, 2, 3$). Graduations generally pass or fail these tests together.

9.7 *The χ^2 test*

The most comprehensive test of the normality of the relative deviations is the familiar χ^2 test. The statistic,

$$\chi^2 = \sum_x (z_x)^2 \tag{9.7.1}$$

is distributed as $\chi^2(N - m)$, where N is the number of ages (or age-groups) and m is the number of parameters in the graduation formula that have been fitted. Computer algorithms are available to calculate $p(\chi^2)$, the probability of a value of χ^2 greater than that observed. Alternatively, if $N - m$ is large enough, say greater than about 30, the probability may be estimated from the fact that the test statistic

$$t(\chi^2) = \sqrt{2\chi^2} - \sqrt{2(N - m) - 1} \tag{9.7.2}$$

has approximately a standard unit normal distribution. If the value of χ^2 is high, so that the value of $p(\chi^2)$ is less than say ·05, and if the graduation has passed the other tests, then it is usually satisfactory.

However, the χ^2 test is as much a test of the hypothesis that the data are independent as a test of the graduation, and in practice many satisfactory graduations show high values of χ^2. This may be accounted for by the presence of duplicates in the data which have not been allowed for. Inspection of the individual values of z_x may show that there is a quite small number of unusually high values; these may be because of duplicates, or they may indicate errors in the data, which should be investigated. The existence of unusually high values of χ^2, with considerable irregularity in the values of z_x, some being high and others low, seems to be a feature of very large investigations where duplicates ought not to occur, such as in graduations of population mortality rates. We cannot account for this observation, but we believe that it is true of other investigations with very large numbers of observations, where apparently significant values of χ^2 are found even when no reason for them should exist.

9.8 *Assessment of battery of tests*

We have now described a battery of statistical tests that can be applied to any one trial graduation. In some cases it will be found that one particular graduation is obviously satisfactory, having passed all the tests, and in other cases a graduation will be found to be obviously unsatisfactory. In practice there are also many intermediate cases, where a graduation passes some tests, but not others. If it fails only the χ^2 test then it is likely that this is because of the data rather than because the graduation is unsatisfactory. A further consideration is whether the shape of the curve of graduated rates outside the main range of the data is sensible. The typical shapes of mortality curves are well known, and a curve that extrapolates in a fairly reasonable way will be considered more satisfactory than one that does not, if extrapolation of the rates to higher or lower ages is desired. But this takes us on to the subject of comparing two different graduations, which is the topic of the next section.

10. CHOICE OF ORDER OF FORMULA

10.1 *Order of Formula*

If one of the families of formulae described in §§4 and 5 is being used for the graduation, it may be necessary to choose which order of formula to use. For example, if the GM(r,s) family is being used, different graduated rates result from different values of r and s. Various tests are available for assisting in this choice, in addition to those described in §9 which test only whether one particular graduation is acceptable, and do not compare two graduations, each of which may be considered acceptable.

10.2 *Likelihood Ratio Test*

First is the likelihood ratio test. Assume that we have fitted a formula with $n = r + s$ parameters, say GM(r,s), or LGM(r,s), and the parameters we have found are the maximum likelihood estimates. We have also fitted a formula with $n + k$ parameters, where one or both of the orders of the polynomials have been increased and neither has been reduced. We can therefore imagine that we are comparing two graduations, each with $n + k$ parameters, with the first graduation having the k extra parameters set to zero. The statistic

$$\begin{aligned} D(k) &= 2\{\log L(n+k) - \log L(n)\} \\ &= 2\log[L(n+k)/L(n)] \qquad (10.2.1) \end{aligned}$$

where $L(n+k)$ is the maximum value of the likelihood for the $(n+k)$-parameter graduation and $L(n)$ is the maximum value of the likelihood for the n-parameter graduation, is distributed as χ^2 with k degrees of freedom (see Kendall & Stuart, 1979). Cox & Oakes (1984) also discuss the application of this test to mortality data. One can therefore test whether the total improvement when k parameters have been added is significant.

10.3 *Akaike Criterion*

If $k = 1$, i.e. one extra parameter is being added, then the statistic $D(1)$ is distributed as χ^2 with one degree of freedom. Since there is approximately a 5% chance that $D(1)$ is greater than 3·8 (i.e. that $\frac{1}{2}D(1)$ is greater than 1·9) and a 95% chance that it is less than this value, we see that one extra parameter can be considered to provide a 'significant' improvement in the graduation if it increases the log-likelihood by 1·9. This is the same test as the 'level of support' described by Edwards (1972) and used by the CMI Committee (1976). If k extra parameters are added in sequence, the log-likelihood needs to be increased by approximately $2k$ for the improvement at *each* step to be considered 'significant'. Note that this is different from what we have described above where we tested whether the k added parameters are significant *in aggregate*. Thus one criterion for the best fitting formula, when at each step the order of one of the polynomials is kept fixed and the order of the other is increased by one, is to choose the order of formula that maximizes the Akaike Criterion

$$\text{AC} = \log L(n + k) - 2k. \qquad (10.3.1)$$

(See Akaike, 1978 and 1985.)

Since the log-likelihood is related to the χ^2 statistic (see § 6.2 above), we can express the same result by noting that a reduction of χ^2 by 3·8 for each extra parameter is worth while. However, the value of χ^2 in mortality experiences is so often confused by the possible presence of duplicates that it is only appropriate to use this criterion rigorously when one can be confident that duplicates are not an issue. Nevertheless the test is extremely useful as a first guide to the order of formula to choose.

10.4 *Maximum* $p(\chi^2)$ *or Minimum* $t(\chi^2)$

A similar test is the maximum $p(\chi^2)$ or its equivalent, the minimum $t(\chi^2)$ test (see McCutcheon, 1984). As the number of parameters is increased the value of χ^2 can be expected to reduce. But the number of degrees of freedom is also reduced. If the number of degrees of freedom, n, is large enough (in practice greater than about 30) then $t(\chi^2)$ is approximately normally distributed with zero mean and unit variance (see § 9.7). Thus, if the number of degrees of freedom is large enough, the value of $t(\chi^2)$ can be used as an alternative statistic to $p(\chi^2)$. As the number of parameters increases, the number of degrees of freedom reduces, the value of χ^2 reduces, and in the first place the value of $t(\chi^2)$ usually reduces, while that of $p(\chi^2)$ usually increases. Beyond a certain point, when the addition of extra parameters has little effect on the graduation, the value of $t(\chi^2)$ is usually found to increase, and that of $p(\chi^2)$ to reduce. Either statistic therefore gives a possible criterion for deciding on a best graduation. Again, the usefulness of this test is diminished by the presence of duplicates, if these have not been allowed for explicitly.

10.5 *Confidence Intervals for Parameters*

The calculation of the information matrix and hence the variance-covariance matrix of the estimates of the parameters in the formula for the appropriate death rates (see § 8) gives yet another guide. In general the vector of estimated parameters, $\hat{\boldsymbol{\alpha}}$, can be assumed to be multivariate normally distributed, given that the true, but unknown parameter vector is $\boldsymbol{\alpha}$. The true value of each parameter is therefore probably (with a roughly 95% probability) within a 2σ confidence interval on each side of the estimated value. If this confidence interval for one particular parameter includes zero, then it is not unreasonable to assume that the true value of that parameter might well be zero, so that the term corresponding to that parameter could be omitted from the formula. In practice we have found that, as the order of a particular formula is increased, the last parameter to be added eventually is not significant in this sense, and we have gone too far. Sometimes we find that, when the order is increased beyond a certain point, several parameters become not significant, even though they were significant with lower orders of formula. It is as if there were only a limited amount of significance to spread around, so that, if too many parameters are chosen the significance is spread too thinly. We do not have a formal statistical explanation of this phenomenon.

11. ESTIMATION OF CONFIDENCE INTERVALS FOR q AND μ

11.1 *Simulation of Parameter Sets*

The variance-covariance matrix of the parameter estimates has allowed us to estimate confidence intervals for the values of the parameters, as we have described in § 10.5 above. It is also possible to use the matrix to estimate confidence intervals for the values of q or μ themselves. In some special cases this may be done analytically, but it is always possible to obtain estimated confidence intervals by Monte-Carlo simulation methods.

It is a property of the method of maximum likelihood that, if the true (but unknown) value of the parameter vector is $\boldsymbol{\alpha}$, the set of maximum likelihood estimates for $\boldsymbol{\alpha}$ derived from a series of experiences, all of which have the same exposures to risk as the single observed experience but different numbers of deaths, is asymptotically multivariate normally distributed with mean value the true vector $\boldsymbol{\alpha}$ and variance-covariance matrix the inverse of the information matrix, evaluated at $\boldsymbol{\alpha}$. Since we do not know $\boldsymbol{\alpha}$ and have in fact only a single observed experience (with the resulting estimate $\hat{\boldsymbol{\alpha}}$), we take $\hat{\boldsymbol{\alpha}}$ as the mean of this distribution and use the corresponding variance-covariance matrix (described in § 9 above) to generate a set of vectors $\{\hat{\boldsymbol{\alpha}}_i : i = 1, 2, \ldots\}$, which we take as an approximation to the maximum likelihood estimates resulting from different sets of observed numbers of deaths $\{A_x\}$.

How this is done is described below. For each sample vector $\hat{\boldsymbol{\alpha}}_i$ we then calculate and record the corresponding values of q or μ over the appropriate range of ages. When a sufficiently large set of vectors $\{\hat{\boldsymbol{\alpha}}_i\}$ has been obtained, the corresponding values of q or μ will—provided the exposures to risk are sufficiently large—be distributed much as the graduated values would be in a set of experiences with the same exposures to risk as the observed experience and the same underlying mortality rates, but with different numbers of deaths. Conversely, the distribution of these rates provides a measure of the possible spread of the true rates of mortality around the values resulting from our maximum likelihood estimate $\hat{\boldsymbol{\alpha}}$.

In order to simulate values from a multivariate normal distribution it is convenient to decompose the parameters into a set of orthogonal (which in this context means independent) random variables. Assume that there are n parameters represented by the column vector $\mathbf{c}$. These have means represented by the column vector $\mathbf{M}$, and a symmetric variance-covariance matrix V. We postulate a set of n independent standard normal random variables represented by the column vector **e**. Suppose that the vector **c** is related to the vector **e** by a transformation $\mathbf{c} = A\mathbf{e} + \mathbf{M}$, where A is an n by n matrix. Then **c** has a mean **M**, and, if A is suitably chosen, variance-covariance matrix V. To show that **c** has mean **M** is trivial: clearly $E[\mathbf{c}] = A \cdot E[\mathbf{e}] + \mathbf{M} = \mathbf{M}$ (since $E[\mathbf{e}] = \mathbf{0}$). The matrix A needs to be chosen so that $E[(\mathbf{c} - \mathbf{M})(\mathbf{c} - \mathbf{M})'] = V$, i.e. $E[A\mathbf{e}\mathbf{e}'A'] = V$. Since $E[\mathbf{e}\mathbf{e}'] = I$, the identity matrix, we just need $AA' = V$.

A can be chosen in many ways, but it is most convenient to choose A as a lower triangular matrix, L. It is well known that, for any non-singular variance-covariance matrix V, there is a unique L such that $LL' = V$. The actual mechanics of calculating $A = L$ are not difficult. Because the terms in A above the diagonal are zero, we have first $a_{11}^2 = v_{11}$ where a_{ij} is the term in the i^{th} row and j^{th} column of A, and v_{ij} is defined likewise for V; so we can calculate a_{11}. In the second row we have $a_{21}a_{11} = v_{21}$ and $a_{21}^2 + a_{22}^2 = v_{22}$, so we get a_{21} and a_{22}. In the next row we derive three terms in succession, and so on. The matrix L is called the Cholesky decomposition of A.

It is not difficult to generate a sequence of pseudo-random unit normal variates, using for example Marsaglia's polar method (see Maturity Guarantees Working Party, 1980, Appendix E, or Press *et al.*, 1986) or the Box–Müller method (Press *et al.*, 1986 or Rubinstein, 1986). We have found the former method to be faster on the computers we have used, but algorithms already programmed for the latter may be more readily available. For the i^{th} vector we generate n such pseudo-random normal variates, which are considered as a column vector $\mathbf{e}_i$. The vector $\hat{\boldsymbol{\alpha}}_i$ is then calculated as $\hat{\boldsymbol{\alpha}}_i = A\mathbf{e}_i + \mathbf{M}$ and sample values of q or μ for a range of ages are calculated using this vector. If the sampling process is repeated say a further 100 times, we shall obtain 101 sets of parameter values in all, and 101 sets of values of q or μ. Sample statistics for each of these variables (parameters and mortality rates) can be calculated, including their means and variances, and the covariance matrix of the sample parameter values. The means of the sample parameters should be close to the means used to generate them, i.e. the maximum likelihood estimates, and the sample variance-covariance matrix should be similar to the matrix V.

11.2 *Standard Errors of the Estimates of Mortality Rates*

The means of the sample mortality rates should be similar to the graduated rates based on the maximum likelihood parameters. There may, however, be some bias in them, depending on the formula used.

For example, suppose that μ_x is given by a Gompertz formula,

$$\mu_x = \exp(\alpha_1 + \alpha_2 x) \qquad (11.2.1)$$

where α_1 and α_2 are bivariate normally distributed, so that (for fixed x) $B = \alpha_1 + \alpha_2 x$ is normally distributed. Let the mean and variance of B be $m(B)$ and $s^2(B)$ respectively. Then μ_x is log-normally distributed with mean

$$E[\mu_x] = \exp\left[m(B) + \frac{s^2(B)}{2}\right] \qquad (11.2.2)$$

and variance

$$\text{Var}\,[\mu_x] \;=\; \exp\,[2m(B) + s^2(B)]\cdot[\exp\,\{s^2(B)\} - 1] \qquad (11.2.3)$$

(See, for example, Hogg & Klugman, 1985.)

In this case the rate based on the parameter mean, viz

$$\exp\,\{m(B)\}$$

is the *median* value of μ_x and is lower than the mean. However, provided that the values of the parameter variances are not too large, in practice the sample values of q or μ are approximately normally distributed, and we can derive a 95% confidence interval for them in the usual way from the sample variances.

11.3 *Quantile Plots*

It is illuminating to plot the sample values of q_x say on a graph. Rather than plot every one of the simulated samples, it is convenient to plot selected quantiles, for example the 1st, 3rd, 5th, 11th, 21st, 51st, 81st, 91st, 97th, 99th and 101st highest, at each age, out of say 101 samples. These typically lie along the line showing the maximum likelihood fit, but with a scatter that indicates the level of confidence one can have in the graduated values.

If the experience is a large one, the confidence intervals both for the parameter estimates and for the graduated rates are small, and the quantile lines lie close together. Towards the ends of the age range in the experience, the quantiles spread out a bit, displaying typically a sheaf-shaped pattern. This indicates that there is greater uncertainty about the graduated rates in this part of the age-range. The sample rates may be carried on beyond any actual data, and this gives an indication of the region within which extrapolated rates might lie, based on the experience from which the rates have been derived. If the experience is a small one, the confidence intervals widen, and the plotted quantiles may cover quite a wide range. Further, if the formula used for the graduation has too many parameters, so that, as noted in § 10.5, the standard errors of the parameter estimates become larger, so also do the confidence intervals for the mortality rates, and sometimes the sheaf-shaped region bursts out erratically, especially at the ends of the range.

Inspection of these quantile plots also helps one to choose between graduations with different formulae or different orders of the same family of formulae. If the sheaf-shaped region is tightly bound in one graduation and bursts out loosely in another, then the former is to be preferred, since the latter is probably an example of over-fitting. However, if the values of the actual (crude) rates are also plotted, one can see whether the sheaf for a chosen graduation satisfactorily covers the actual rates. If one or two points representing the actual rates appear irregularly on either side of the sheaf, then the graduation is reasonably good. Usually these outlying points indicate high (absolute) values of the relative deviation z_x. If a string of points representing the actual rates appears to one side or another of the sheaf, then the trial graduation does not run down the middle of the data satisfactorily, and a higher order of formula, or a formula from a different family should be tried. One can say as a general rule that the tightest sheaf that covers the data is the best, but this generalization does not necessarily cover all circumstances, and it is in any case a matter of judgement by eye as to which of two similar graduations is the tighter or covers the data better.

11.4 *Distribution of log-likelihood*

At the same time as calculating the set of mortality rates resulting from each sample vector $\hat{\boldsymbol{\alpha}}_i$ we calculate $L(\hat{\boldsymbol{\alpha}}_i)$, the corresponding likelihood for the observed experience. In each case the value of $L(\hat{\boldsymbol{\alpha}}_i)$ must be less than $L(\hat{\boldsymbol{\alpha}})$, the maximum value. The likelihood ratio test (see § 10.2 above) shows that asymptotically $-2 \log (L(\hat{\boldsymbol{\alpha}}_i)/L(\hat{\boldsymbol{\alpha}})) = 2(\log L(\hat{\boldsymbol{\alpha}}) - \log L(\hat{\boldsymbol{\alpha}}_i))$ is distributed as $\chi^2(n)$, where n is the dimension of the vector $\boldsymbol{\alpha}$.

12. RECONSIDERATION OF THE DATA

12.1 *Assessment of the Test Results*

It is now appropriate to review what has been written above about choosing which out of many possible graduations is the most satisfactory, and to bring out other considerations that may be relevant. We start by assuming that a particular family of formulae, for example the GM(r, s) family, has been chosen. If we are dealing with mortality data at adult ages, then a Gompertz formula GM(0, 2) is the lowest order than can possibly be suitable, and it is often convenient to start there. (To put $s = 1$ is ambiguous, since the exponential term is then confounded with the constant term in the polynomial in x of order r; to put $s = 0$ is possible, but leaves the formula just as a polynomial in x, which has in practice not been found to fit mortality rates well.) In some cases, particularly if the experience is not very large and is confined to ages above about 30, we may find that we need not go any further.

However, it seems to us desirable always to investigate the effect of adding one parameter too many. Starting with GM(0, 2) we could add one parameter to the first polynomial, giving GM(1, 2) (Makeham's formula), or one to the second polynomial, giving GM(0, 3). If we find that one or other of these is better than GM(0, 2), we may wish to go further, trying GM(2, 2), GM(1, 3) or GM(0, 4). We can carry on, increasing the total number of parameters by one each time, and trying all the possible formulae in the family with that number of parameters. One line of approach is to choose the best of the n-parameter formulae, using for example the criterion of the maximum value of the log-likelihood, and advancing forward by adding one parameter to each of those polynomials in turn; whether this can readily be done depends on the cost of computing, but we have not found it expensive to try a wide range of orders in any one sequence of trials. If the best of the $(n + 1)$-parameter formulae has a log-likelihood no more than 2 larger than the best of the n-parameter ones, and if the latter is a special case of the former, then it is probably not worth going further. If the family of formulae is in general a suitable one, then this probably identifies the most suitable graduation. But sometimes the log-likelihood test (or the minimum $t(\chi^2)$ test) draws one on to a higher order formula than seems to us to be justified. This appears to occur when insufficient or no allowance for duplicates has been made, since the value of χ^2 (and the log-likelihood) should have been scaled down to allow for them.

If a graduation of a lower order fails some of the elementary tests, such as the signs test, runs test, serial correlation test, or Kolmogorov–Smirnov test, and a graduation with a higher order passes them, then the latter is to be preferred. But sometimes no graduation of a reasonably low order satisfies these tests. In this case another family of formulae may be tried. There is no point in increasing the number of parameters excessively. It is usually found that the standard errors of the parameter estimates increase, so that many of the parameters appear to be not significantly different from zero. The quantile sheaf also tends to burst out, if too many parameters are included. If the standard errors of the sample mortality rates are calculated, the graduation with the lowest values of these standard errors may be preferred, provided that it satisfies adequately the elementary tests.

12.2 *Visual Inspection*

It is advisable also to look at the chart showing the actual rates and the gates described in § 2.6. If a graduated rate for one age lies outside the corresponding 95% gate, then its relative error, z_x, is outside the range $\pm 1{\cdot}96$. If too many gates are missed, particularly if they are in a sequence of ages, then a higher order formula (or another family) may help. But if the gates at one or two ages are so positioned that no reasonable curve could pass through them without missing some of the neighbouring gates, then one should first suspect erroneous figures, or faulty data, and then probably reach the conclusion that the data itself has caused the problem. There may be unidentified duplicates (as was recorded by the CMI Committee (1976) in relation to the 1967-70 male immediate annuitants data), or there may be some other peculiarity about the data at that age.

In the CMI data relating to male pensioners, retiring at or after the normal age for their scheme, there have always been one or two such ages in the range from 64 to 67 (but not always at the same ages in each quadrennium). Since members of pension schemes typically retire at around those ages, the exposed to risk is changing rapidly, and it is possible that the method used to calculate the exposed to risk is not sufficiently accurate; another possible explanation is that the level of mortality of those who have just retired is unusual; yet another is that if a member dies at around retirement age there is bias as to whether or not he is deemed by the scheme administrators to have gone on pension. None of these explanations is wholly plausible, but it would not be appropriate when graduating an aggregate table like this one to take too much account of this irregularity, even though one or two gates are missed, and one or two ages show unusually high values of z_x.

Sometimes one finds that the data at high ages is particularly erratic, and successive gates are well out of line. This may be because of misstatements of age (in population data) or because unreported deaths are included in the exposed to risk (in life assurance data). In either case it may be more satisfactory to restrict the age range for the graduation, stopping say at age 89, and to allow the graduation to produce extrapolated rates. The CMI Committee (1974) did this when graduating the experience on which the A1967-70 ultimate rates were based.

On the other hand, there is no justification for missing out ages where the data is just scanty, so that the gates are wide and neighbouring gates overlap to a great extent. The crude rates may be erratic, but they do not carry a great weight in the graduation, and the method of grouping described in § 9.2 ensures that they do not upset the tests.

Inspection of a chart on which the full range of graduated rates, including extrapolations beyond the age range of the data, is desirable. In some cases extrapolation of the rates may not be needed, but it is frequently necessary to continue the rates to high ages to complete a mortality table. If one graduation produces sensible extrapolations, and another (equally satisfactory in other respects) does not, then the former may be preferred, since it would then be possible to use the graduated rates without special adjustment. In other circumstances it may in any case be necessary to extend the rates on the basis of other data, so this consideration may not apply. Formulae with too high orders of polynomial may fit the data in the main age range well, but they have a habit of turning in an undesirable way beyond this range. It seems implausible for mortality rates to reduce steeply at the highest ages, and a formula that has a pronounced turning point in the useable age range is undesirable (though it may not matter if the turning point is at a sufficiently high age).

12.3 *Related Experiences*

If the experience being graduated is one of a related set of experiences, such as the data for successive durations in a select investigation, then a desirable feature of the graduated rates may be that they run consistently. If it is clear that the level of mortality at the main age ranges increases with duration, then a set of graduated rates that run suitably 'parallel', i.e. do not overlap in the desired age range, may be preferable to ones that cross over implausibly. For this sort of reason a graduation that has a generally desirable shape may well be more suitable than another that has a less satisfactory shape, even though the latter might be preferred on the basis of statistical tests alone. The battery of statistical tests that we have described are a guide to assist the actuary's judgement, not a set of rules to override it. The examples we discuss in §§ 15 to 17 may help to illustrate this.

13. COMPARISON OF TWO EXPERIENCES

13.1 *Comparison of Crude Rates*

We may wish to compare the data from two different mortality experiences to see whether they can be considered to have experienced the same levels of mortality. The experiences may be for two different durations in the same investigation, or for two different periods for the same investigation, or for two different investigations. Whatever the contrast, the comparison can be done in two ways: first by comparing the data from the two experiences directly, perhaps before graduation, and secondly by graduating the data for the two experiences separately and comparing the graduated rates. Different tests apply in these two circumstances. We discuss them in turn.

Assume that we have data for two experiences, denoted by I and II. For each we have a sequence of values of exposed to risk and of actual deaths, for a range of ages. The ranges may not be identical for the two experiences, but we assume that there is a reasonable overlap of ages. (Otherwise there is not much to compare!) We may wish to restrict our comparison to roughly matching ages. We may wish first to adjust the exposed to risk and actual deaths in either experience or both, so as to eliminate the effect of duplicates. This can be done by dividing the exposed to risk and the number of deaths at each age by the appropriate variance ratio, as described in § 3.2. We then suggest grouping the ages, perhaps in the way described in § 9.2, so that the number of deaths (after adjustment) in each experience in each group is at least some number, such as 5. We can then assume at least approximate symmetry and probably approximate normality in the distributions of actual deaths in the two experiences.

We denote the numbers of exposed to risk at age x (or in age-group x) in the two experiences by R_x^{I} and R_x^{II}, and the numbers of actual deaths by A_x^{I} and A_x^{II}, in each case after adjustment by any variance ratio (if appropriate). We first calculate the crude mortality rates at each age for the two experiences. Assume that we have central exposed to risk, that we wish to work with μ, and that the quotient A_x/R_x gives the crude value of the force of mortality at exact age x. (If we use q, the methods are exactly comparable.) We denote the crude rates as

$$\mu_x^{\mathrm{I}} = A_x^{\mathrm{I}}/R_x^{\mathrm{I}} \qquad (13.1.1)$$

and

$$\mu_x^{\mathrm{II}} = A_x^{\mathrm{II}}/R_x^{\mathrm{II}} \qquad (13.1.2)$$

We shall wish to use also the pooled data, and we shall denote the corresponding exposed to risk, actual deaths, and crude rates for this by

$$R_x = R_x^{\mathrm{I}} + R_x^{\mathrm{II}} \qquad (13.1.3)$$

$$A_x = A_x^{\mathrm{I}} + A_x^{\mathrm{II}} \qquad (13.1.4)$$

and

$$\mu_x = A_x/R_x \tag{13.1.5}$$

13.2 *Comparison of Levels of Mortality*

We first wish to test whether the general level of mortality rates is the same for each experience. (If it is not, we may decide to go no further.) We use the signs test for this, calculating the signs of the differences $\mu_x^{\text{I}} - \mu_x^{\text{II}}$. On the basis of the null hypothesis that the true mortality rates for the two investigations are the same, at each age the difference between the crude rates in the two populations is equally likely to be positive or negative. (At least for practical purposes this may be considered the case; strictly it is true only asymptotically as the exposures to risk increase.) Accordingly the number of positive differences is distributed binomially $B(N, 1/2)$, where N is the number of ages (or age-groups) that we are comparing.

13.3 *Comparison of Shapes*

Even if the signs test does not show a significant difference in the overall level of mortality in the two experiences, the levels may not be 'parallel', so that one experience shows higher mortality at some ages and lower at others. If the mortality of the two experiences is the same, then the differences should not only be equally likely to be positive and negative, but the pattern of positive and negative signs should be random. The runs test can be used on the signs of the differences of the crude rates.

13.4 *The χ^2 Test*

We can also use the χ^2 test to compare the distribution of the rates. To do this we use the pooled experience to estimate μ_x. We then calculate the expected deaths at each age for each experience as

$$E_x^{\text{I}} = R_x^{\text{I}}\mu_x \tag{13.4.1}$$

and

$$E_x^{\text{II}} = R_x^{\text{II}}\mu_x \tag{13.4.2}$$

We also calculate the deviations and relative deviations at each age for each experience as

$$\text{Dev}_X^{\text{I}} = A_x^{\text{I}} - E_x^{\text{I}} \tag{13.4.3}$$

$$z_x^{\text{I}} = \text{Dev}_x^{\text{I}}/\{E_x^{\text{I}}\}^{1/2} \tag{13.4.4}$$

$$\text{Dev}_x^{\text{II}} = A_x^{\text{II}} - E_x^{\text{II}} \tag{13.4.5}$$

and

$$z_x^{\text{II}} = \text{Dev}_x^{\text{II}}/\{E_x^{\text{II}}\}^{1/2} \tag{13.4.6}$$

Finally we calculate χ^2 as

$$\chi^2 = \Sigma[(z_x^{\mathrm{I}})^2 + (z_x^{\mathrm{II}})^2] \tag{13.4.7}$$

and note that χ^2 is distributed as $\chi^2(N)$. (Although it is the sum of $2N$ terms, N degrees of freedom are used up in the calculation of the pooled values of μ_x.)

A sufficiently low value of χ^2 indicates that the experiences have similar patterns of mortality. A too high value does not necessarily indicate the opposite. Inspection of the individual values of z_x^{I} and z_x^{II} may show that at one or two ages the data are more likely to be the problem.

If the two experiences clearly do not have significantly different crude rates, then there is little justification for graduating them separately. The tests we shall describe below would probably lead us to the same conclusion if we were to graduate them separately. At the other extreme, if the levels of mortality, as shown by the signs test, are clearly different, or if the shapes, as shown by the runs test, are clearly different, then there is little justification for graduating the pooled experience.

However, if the levels and shapes of the crude rates are similar and the χ^2 test alone shows a difference, then we would prefer to graduate the two experiences separately, and test the graduated rates, as described below, with a view to pooling the experiences, and graduating the pooled rates, if the graduations are not significantly different. In some cases a high value of χ^2 in the test of the crude rates corresponds to a high value of χ^2 for one or other (or both) of the experiences when each is compared with its own graduation. Often the same age or ages produce high values of z_x. This is not in itself a reason for not pooling. In our view, pooling is to be preferred unless it cannot possibly be justified. There is seldom a good reason for making distinctions between durations or experiences for their own sake, but there is often a temptation to make spurious distinctions where none exist.

13.5 *Comparison of Graduated Rates*

We now describe the test that might be applied to compare the graduated rates of two experiences. We assume that we find that the same formula fits the two experiences, if necessary by increasing the order of the polynomials in the formulae to make them correspond. We assume that this formula requires n parameters. We denote by $\hat{\boldsymbol{\alpha}}^{\mathrm{I}}$ and $\hat{\boldsymbol{\alpha}}^{\mathrm{II}}$ the parameter vectors obtained (by maximum likelihood or minimum-χ^2 methods) for the graduations of the two experiences. We denote by V^{I} and V^{II} the corresponding variance–covariance matrices. We then define the measure of 'distance' between the two sets of graduated parameters as

$$D = (\hat{\boldsymbol{\alpha}}^{\mathrm{I}} - \hat{\boldsymbol{\alpha}}^{\mathrm{II}})'(V^{\mathrm{I}} + V^{\mathrm{II}})^{-1}(\hat{\boldsymbol{\alpha}}^{\mathrm{I}} - \hat{\boldsymbol{\alpha}}^{\mathrm{II}}) \tag{13.5.1}$$

If experiences I and II are two samples from the same underlying experience, then D is distributed as $\chi^2(n)$. We can thus use a χ^2 test on the distance between the two experiences in the n-dimensional parameter space. This gives a simple and conclusive test for the comparison of two graduations.

14. CONSTRUCTION OF THE MORTALITY TABLE FROM THE GRADUATED RATES

14.1 Suppose first that the graduation has produced a formula for q_x. In this case, since

$$l_{x+1} = (1 - q_x)l_x \tag{14.1.1}$$

the values of l_x in the resulting mortality table are readily obtained successively starting from any chosen radix.

If the graduating function is valid only over a restricted range of ages, in certain circumstances it may be necessary to extrapolate the mortality rates to a wider age-range. (For example, in an investigation into the mortality of assured lives it may be desirable for practical purposes to extend the table 'downwards' to age 0.) In such a situation one must consider whether or not the graduating formula may be used for ages outwith the basic age-range. If the formula is considered inappropriate at these ages, *ad hoc* extrapolation methods will have to be used (see, for example, CMI Committee, 1974 and 1976).

It will generally be useful to include in the published table values of μ_x at each age. Such values can be estimated from the sequence of values $\{l_x\}$ in several ways. For example, one might assume that over a series of appropriate (short) age-ranges the force of mortality (as a function of age) is a polynomial of low degree. The resulting set of values $\{\mu_x\}$ will be sufficiently accurate for all practical purposes and may, therefore, be included in the published table. (It is perhaps worth noting that an alternative method of estimating the force of mortality, based on the assumption that over short age-ranges l_x itself is a polynomial of low degree is, for theoretical reasons, a less desirable procedure (see McCutcheon, 1983).) Similar methods may also be used for a select table.

14.2 Suppose, alternatively, that the graduation has produced a formula for the force of mortality μ_x. In this situation the sequence of values $\{q_x\}$ is required, in order to construct the mortality table.

If the graduating function can be integrated exactly, the value of q_x at each age is easily calculated as

$$q_x = 1 - \exp\left\{-\int_x^{x+1} \mu_y \mathrm{d}y\right\} \tag{14.2.1}$$

If exact integration is not possible, since the force of mortality can be evaluated by formula at *any* age (and not only at integer ages) it is a trivial matter to evaluate the right-hand side of equation (14.2.1) to any desired degree of accuracy by standard methods of approximate integration (see Conte & de Boor, 1980). Even using only the values of μ_x at integer ages, one can generally estimate the value of the integral in equation (14.2.1) sufficiently accurately for most practical purposes.

In certain situations one of the simple approximations

$$q_x \simeq 1 - \exp\{-\mu_{x+1/2}\}, \tag{14.2.2}$$

$$q_x \simeq 1 - \exp\{-\tfrac{1}{2}(\mu_x + \mu_{x+1})\} \tag{14.2.3}$$

or

$$q_x \simeq \frac{\tfrac{1}{2}(\mu_x + \mu_{x+1})}{1 + \tfrac{1}{2}\mu_{x+1}} \tag{14.2.4}$$

might be appropriate. (See Waters & Wilkie, 1987.) In general, however, greater accuracy than that provided by these first approximations is desirable.

14.3 If the graduation has produced a formula for the central death rate m_x, it is necessary to obtain values of both q_x and μ_x at each age.

Provided that the formula for m_x is valid at all ages above a certain value (β say) and the graduating function can be easily integrated, the values of q_x and μ_x at each age greater than β can be found exactly quite simply by letting

$$g_x = m_x \exp\left\{-\int_\beta^x m_y \mathrm{d}y\right\} \tag{14.3.1}$$

and

$$h_x = -g'_x = [(m_x)^2 - m'_x] \exp\left\{-\int_\beta^x m_y \mathrm{d}y\right\} \tag{14.3.2}$$

Then (subject to convergence conditions which will be satisfied in most practical situations) we have, for $x \geqslant \beta$,

$$q_x = \frac{g_x}{\sum_{r=0}^{\infty} g_{x+r}} \tag{14.3.3}$$

and

$$\mu_x = \frac{\sum_{r=0}^{\infty} h_{x+r}}{\sum_{r=0}^{\infty} g_{x+r}} \tag{14.3.4}$$

(see McCutcheon, 1971). Although these last two equations involve infinite series, their evaluation by computer to any desired degree of accuracy is a trivial matter. Having determined the two sequences $\{g_x\}$ and $\{h_x\}$ ($x = \beta, \beta + 1, \beta + 2, \ldots$), we can immediately construct the sequences $\{q_x\}$ and $\{\mu_x\}$ at these same ages from equations (14.3.3) and (14.3.4).

If it is impractical to integrate the graduating function for m_x or to carry out the above procedures, alternative methods are readily available. For example, at each age we may estimate q_x either in terms of m_{x-1} and m_x or in terms of m_x and m_{x+1} by simple quadratic-based approximations. (See McCutcheon, 1977.)

Then, using the values obtained for the sequence$\{q_x\}$, we can estimate the force of mortality at each age as in § 14.1 above.

Alternatively, except at high ages, we might use the well-known approximation

$$\mu_x \simeq m_{x-1/2} \tag{14.3.5}$$

to determine the force of mortality at each integer age, the right-hand side of equation (14.3.5) being evaluated by formula. Then, using these values of the force of mortality, we may estimate q_x by an appropriate approximate integration procedure (as in § 14.2 above).

15. EXAMPLE I—PENSIONERS' WIDOWS

15.1 *Introduction*

We now show how the principles described above can be applied by giving practical examples of the graduation of particular experiences. Each experience is one of those whose graduations are discussed in the accompanying report, and each relates to one of the CMI investigations for the period 1979 to 1982. Our first example is a relatively new investigation, that of the widows of life office pensioners. It has relatively few deaths, 692 in all, and it turns out to be rather easy to find a satisfactory graduation, since a simple formula with two parameters suffices. Consequently, its simplicity makes it suitable for exemplifying the various alternative methods and formulae described above.

Our second example (in § 16) uses a much larger experience, that for male life office pensioners retiring at or after their normal retirement age. This has 85,426 deaths, but it too is a more simple example, because of its restricted age range, absence of select period, and apparent absence of duplicates. It has, however, data both on a 'Lives' and on an 'Amounts' basis.

Our third example (in § 17) is more complex, and it includes a number of the complications absent from our second example. This is for male permanent assured lives (UK) with a total of 95,023 deaths spread over six durations, a full age range and a considerable number of duplicates.

In order to make comparisons between different GM and LGM formulae more meaningful, in the remainder of this paper we adopt the convention that the coefficients of the first polynomial in a GM(r, s) formula are denoted, by a_0, $a_1, \ldots .a_{r-1}$. We call these the 'r-coefficients'. If $r = 0$, there are no r-coefficients. Similarly we denote by $b_0, \ldots .b_{s-1}$ the coefficients of the polynomial in the exponential term and refer to these as the 's-coefficients'. If $s = 0$, there are no s-coefficients. Throughout we use Chebycheff polynomials of the first type (see § 4.2 above), denoting by $C_i(t)$ the polynomial in t of degree i, and work with the scaled variable

$$t = \left(\frac{x - 70}{50}\right)$$

These conventions being clearly understood, we may then write

$$\mathrm{GM}_{\alpha}^{r,s}(x) = \left\{a_0 + a_1 C_1\left(\frac{x-70}{50}\right) + \ldots + a_{r-1}C_{r-1}\left(\frac{x-70}{50}\right)\right\} + \exp\left\{b_0 + b_1 C_1\left(\frac{x-70}{50}\right) + \ldots + b_{s-1}C_{s-1}\left(\frac{x-70}{50}\right)\right\} \quad (15.1.1)$$

where, as before, if $r = 0$, the right-hand side of this last equation is to be interpreted as containing only the exponential term and, if $s = 0$, it is to be interpreted as containing only the polynomial term.

This notation is used also for LGM(r, s) formulae, $\mathrm{LGM}_{\alpha}^{r,s}(x)$ being defined by equations (4.1.6) and 15.1.1).

15.2 *Widows of life office pensioners*

We consider first the experience of the widows of pensioners of schemes insured by life offices. This investigation began only in 1975 and this is the first occasion that the experience of spouses has been considered by the CMI Committee for graduation. We therefore had no prior experience to guide or to hinder us. The total central exposed to risk for this experience was 28,386.5 years (initial 28,732.5), and there were 692 deaths. The numbers at each age appear in Tables 15.5 (central) and 15.6 (initial). The extreme limits of age observed in the exposed to risk were 17 to 108 (nearest birthday), but no deaths were observed below age 45 or above age 98. Outside the range of ages from 38 to 85 inclusive the central exposed to risk at each age was less than 100, and outside the range of ages from 59 to 88 inclusive the number of deaths at each age was less than 10, so the bulk of the deaths were concentrated into about a 30-year age span.

There was no reason to suppose that a large number of duplicates existed in this experience, and no information from which variance ratios could be derived, so it was assumed that each observation represented one 'life'. The subsequent results did not suggest that this assumption might have been unjustified.

The experience was graduated both by using 'initial' exposures and graduating q_x, which is the traditional way used for CMI graduations, and by using 'central' exposures and graduating μ_x, which is possibly preferable. We also graduated q_x with 'central' exposures, but to save space we do not report on the results here. Since the CMI Bureau calculates initial exposures by adding half the number of observed deaths to the mean of the census populations for the in-force, it is an easy matter to subtract half the deaths from the initial exposures to obtain the central exposures.

The crude rates of μ_x or q_x for each age were calculated, and those for μ_x are plotted in Figure 15.1. Confidence intervals for μ_x and q_x were also calculated, using the methods described in § 2.6, and using the exact Poisson or Binomial method for each age. The maximum number of deaths at any one age was 33 (at age 75), and our criterion has been to use the normal approximation only when the number of actual deaths exceeds 60. The upper and lower limits of the 95% confidence intervals for μ_x are also shown in Figure 15.1 (along with a possible graduation formula, GM(0, 2), discussed below). The corresponding figure for q_x has a very similar appearance.

It can be seen that the confidence intervals are extremely wide at young ages, where there were no deaths, and reduce as the exposed to risk increases. Even from age 45 to about 60 the observed rates are erratic and the confidence intervals are wide. From ages 60 to 90 or so the observed rates rise in the usual way of almost all mortality experiences, and the confidence intervals are each of a similar width. Above age 90 the crude rates become more erratic and the confidence intervals widen again, extending almost the whole width of the interval (0, 1) in the ages above 95.

It is pretty clear in general where the graduated rates should lie, and pretty clear that a straight line drawn on the semi-logarithmic graph would pass through most of the gates. A straight line corresponds to a Gompertz or GM(0, 2) formula, and this is clearly the starting point of our investigations. We shall pre-empt our conclusions by saying now that this formula (and the related logistic LGM(0, 2) formula) proved to be our finishing point too, but discussion of the appropriate order of formula to use is deferred to § 15.3.

For each of the functions

(A) μ_x and

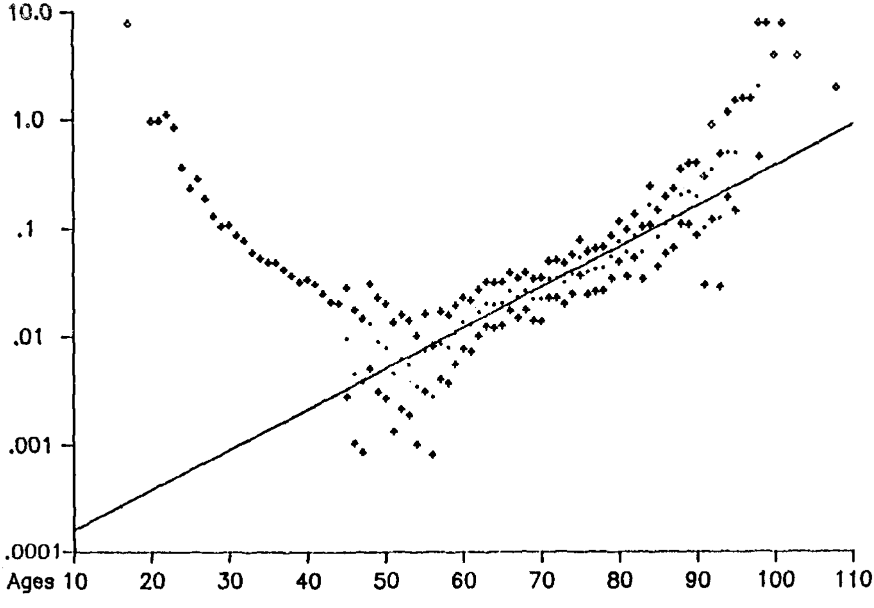

Figure 15.1. Widows of pensioners crude rates and gates: μ_x = GM(0, 2).

(B) q_x with initial exposures,
we fitted the formulae
(a) Gompertz GM(0, 2) and
(b) logistic LGM(0, 2),
using in each case the maximum value of
(1) the likelihood (L_1),
(2) the normal approximation to the likelihood (L_2), and
(3) $-\frac{1}{2}\chi^2$ (L_3).
This gives $2 \times 2 \times 3 = 12$ different graduations, which we denote for reference purposes by Aa1, Bb2, etc.

(In fact, because the available computer routines were *minimization* routines, in each case we calculated the optimum parameters by minimizing $-L_1$, $-L_2$, or $-L_3$ as appropriate. We sometimes used two different minimization methods (steepest descent and the simplex method), which gave the same answers (a comforting discovery).) We calculated the statistics for the various tests described in § 9, and also the information matrix, H, at the optimal point. From the inverse of H we calculated the standard errors of the parameter estimates (see § 8 above).

In graduations of μ_x we calculated the corresponding values of q_x at integral ages by approximate integration, using Simpson's rule repeatedly until successive answers were near enough the same. In graduations of q_x we could calculate q_x at integral ages from the appropriate formula. We used the estimated optimal parameters and the derived variance-covariance matrix to simulate at random 100 sets of parameters and values of μ_x or q_x at each age. In the case of μ_x we then derived simulated values of q_x for integral ages by the approximate formula

$$q_x = 1 - \exp(-\mu_{x+1/2})$$

and calculated standard errors for q_x from the observed simulated values of q_x.

The results of each of the graduations are summarized in Tables 15.1 to 15.4. Each table is for one combination of function and formula, and shows results for all three criteria. Thus Table 15.1 shows results for μ_x, formula GM(0, 2), for criteria L_1, L_2 and L_3. The Tables show

Table 15.1	Graduation Aa	μ_x	GM(0, 2)
Table 15.2	Graduation Ab	μ_x	LGM(0, 2)
Table 15.3	Graduation Ba	q_x (initial)	GM(0, 2)
Table 15.4	Graduation Bb	q_x (initial)	LGM(0, 2)

We can first observe that all twelve graduations shown fit the data quite satisfactorily. The p-values for the signs test, runs test and χ^2 test are all satisfactorily large (i.e. well above ·05, the value which would imply a significant deviation from a satisfactory fit). The absolute values of the t-ratios for the serial correlation test are all well below 1·96, the value which, at the 5% level, would imply a significant deviation of the serial correlation coefficient from zero. (T-ratios here, and when the parameter estimates are being discussed, are simply the estimated value divided by the standard error of the estimate; for the serial correlation coefficients the standard errors are $N^{-1/2}$, where N is the number of values of z_x used to calculate the coefficients.) The values of p(KS) are generally well above ·9, though in a few cases they are below this.

The value of $A - E$ (the difference between the actual and expected deaths) is always small, but in two cases it is almost exactly zero; these are graduations Aa1(μ_x), formula GM(0, 2), maximum likelihood) and Bb1 (q_x), formula LGM(0, 2), maximum likelihood), which are the two cases where this result can be expected on theoretical grounds.

The values of the parameters b_0 and b_1 are in all cases very significantly different from zero, and are reasonably similar for all the graduations of any one function and any one formula. But they are by no means the same. The normal approximation does not necessarily produce values any closer to the maximum likelihood values than the cruder L_3 approximation.

The values of q_x produced by the different methods show quite a large range. At age 70, where the standard error of q_x is at a minimum, the range of values is from ·028285 (Ba2) to ·031079 (Ab3), a range of about 10%. This should be compared with the standard errors of the estimates of q_{70} which are around 3·5% to 4% of the value of q_{70}. Different formulae and criteria produce results that appear significantly different from each other. The divergence is even wider at the extreme ages for which values of μ_x and q_x are shown, where the formulae are projected well beyond the range of any significant volume of data.

There is less difference, however, between the values of q_x for our two preferred methods, Aa1 and Bb1. These show values of q_{70} of ·029468 and ·029629, respectively, a difference of only about ·5%, well within one standard error. The differences between the values of q_x for these two graduations remain small relative to their standard errors for all ages from 20 to 100, and exceed one standard error only at age 110.

The graph of the GM(0, 2) formula for μ_x, using the L_1 criterion, is shown on Figure 15.1. It can be seen how it passes through almost all the gates, the only exceptions being at the extreme ends of the data where the numbers of deaths are small.

The results of the simulation of values of μ_x derived from the GM(0, 2) graduation with the L_1 criterion are depicted in Figure 15.2, which shows the crude rates, the graduated rates, and the following quantiles from 100 simulations: numbers 1, 3, 5, 10, 20, 81, 91, 96, 98 and 100. Thus the lowest (dotted) line joins the points representing the lowest values of μ_x simulated at each separate age. These are not necessarily all from one set of simulated parameter values. The 'sheaf' shape can easily be seen.

The standard errors of the estimates of q_x indicate the accuracy which it is reasonable to expect any graduation to achieve. The values of the standard errors for each age are reasonably similar for all the graduations. The value of one standard error is less than 10% of the estimated value of q_x for all ages from about 50 to 90. This is a little wider than the range of ages at which the numbers of deaths at each age exceed 10.

Tables 15.5 and 15.6 give detailed results for graduations Aa1 and Bb1, showing for each age the exposed to risk, R_x, the number of deaths, A_x, the

Table 15.1. *Pensioners' widows, 1979-82. Statistics for graduations Aa. Function: μ_x, formula: GM(0, 2)*

Method Optimization criterion	L_1 Maximum likelihood	L_2 Normal approximation	L_3 $-\frac{1}{2}\chi^2$
Values of criteria at optimum point:			
L_1	− 3003·23	− 3004·86	− 3003·85
L_2	153·61	155·55	152·73
L_3	− 30·24	− 32·40	− 29·60
Values of parameters at optimum point:			
b_0	− 3·553013	− 3·587134	− 3·512447
(standard error)	·039234	·037967	·036668
T-ratio	− 90·56	− 94·48	− 95·79
b_1	4·316579	4·664277	4·343006
(standard error)	·196615	·162352	·159236
T-ratio	21·95	28·73	27·27
Comparison of total actual deaths (A) and total expected (E):			
Total A-E	·00	10·10	− 29·60
Ratio 100A/E	100·00	101·48	95·90
(Ages grouped so that each $E_x \geqslant 5$)			
Signs test:			
Number of +	19	23	17
Number of −	22	18	24
p-(pos)	·3776	·8256	·1744
Runs test:			
Number of runs	21	20	19
p(runs)	·5124	·4120	·3233
Kolmogorov–Smirnov test:			
Max deviation	·0228	·0467	·0245
p(KS)	·9938	·4420	·9839
Serial Correlation test:			
r_1	− ·0747	·0241	− ·0027
T-ratio	− ·48	·15	− ·02
r_2	·1258	·1405	·1163
T-ratio	·81	·90	·74
r_3	− ·0734	·0188	− ·0960
T-ratio	− ·47	·12	− ·61
χ^2 test			
χ^2	38·29	38.97	35·68
Degrees of freedom	39	39	39
$p(\chi^2)$	·5019	·4712	·6223
Specimen values of q_x for integral ages, simulated standard errors, and standard errors as percentage of value of q_x:			
Age 20	·000399	·000273	·000405
(standard error)	·000092	·000049	·000067
percentage s.e.	22·98	18·06	16·52

Table 15.1 (*cont.*)

Method Optimization criterion	L_1 Maximum likelihood	L_2 Normal approximation	L_3 $-\frac{1}{2}\chi^2$
Age 30	·000946	·000695	·000965
(standard error)	·000174	·000102	·000129
percentage s.e.	18·44	14·64	13·33
Age 40	·002242	·001765	·002298
(standard error)	·000316	·000200	·000235
percentage s.e.	14·09	11·32	10·23
Age 50	·005306	·004480	·005469
(standard error)	·000529	·000364	·000398
percentage s.e.	9·97	8·13	7·28
Age 60	·012536	·011348	·012987
(standard error)	·000785	·000598	·000616
percentage s.e.	6·26	5·27	4·74
Age 70	·029468	·028593	·030677
(standard error)	·001176	·001013	·001080
percentage s.e.	3·99	3·54	3·52
Age 80	·068462	·071082	·071575
(standard error)	·003558	·003143	·003334
percentage s.e.	5·20	4·42	4·66
Age 90	·154772	·170902	·162236
(standard error)	·012490	·011301	·010966
percentage s.e.	8·07	6·61	6·76
Age 100	·328796	·378966	·344225
(standard error)	·033918	·030691	·028496
percentage s.e.	10·32	8·10	8·28
Age 110	·611429	·702045	·634218
(standard error)	·061060	·048323	·048582
percentage s.e.	9·99	6·88	7·66

graduated death rate (μ_x or $q_{x-1/2}$, as appropriate) for the formula, the corresponding expected number of deaths, E_x ($R_x\mu_x$ or $R_xq_{x-1/2}$, as appropriate), and the deviation $\text{Dev}_x = A_x - E_x$. Note that the exposed to risk and actual deaths are those for 'age nearest', so the age adjustment b (see § 3.1) is $-1/2$, and the values given are those for μ_x and $q_{x-1/2}$.

Neighbouring ages are then grouped so that the total of E_x in each group is at least 5. The following items are shown for each resulting age group: totals of R_x, A_x, E_x and Dev_x; the standard deviation $(V_x)^{1/2}$ (see § 9.1), the statistic z_x, and the ratio A_x/E_x as a percentage. The values of z_x can be compared mentally with a unit normal variate. It can be seen that only one value in each table, that for age 88, exceeds 2·0 in absolute value. The signs test, and runs test are carried out on the signs of the z_x and the serial correlation test uses the values of the z_x. The sum of the squares of z_x gives the value of χ^2.

Table 15.2. *Pensioners' widows 1979-82. Statistics for graduations Ab. Function: μ_x, formula: LGM(0, 2)*

Method Optimization criterion	L_1 Maximum likelihood	L_2 Normal approximation	L_3 $-\frac{1}{2}\chi^2$
L_1	− 3003·17	− 3004·90	− 3003·78
L_2	154·98	157·12	153·93
L_3	− 31·87	− 34·42	− 31·23
Values of parameters at optimum point:			
b_0	− 3·512845	− 3·542254	− 3·469051
(standard error)	·040636	·039457	·038181
T-ratio	− 86·45	− 89·78	− 90·86
b_1	4·526366	4·923193	4·532029
(standard error)	·215332	·177519	·171787
T-ratio	21·02	27·73	26·38
Comparison of total actual deaths (A) and total expected (E):			
Total A-E	·34	7·09	− 29·23
Ratio 100A/E	− 100·05	101·03	95·95
(Ages grouped so that each $E_x \geqslant 5$)			
Signs test:			
Number of +	19	22	17
Number of −	21	18	23
p(pos)	·4373	·7852	·2148
Runs test:			
Number of runs	18	18	22
p(runs)	·2170	·2293	·7377
Kolmogorov–Smirnov test:			
Max deviation	·0267	·0513	·0270
p(KS)	·9658	.3254	·9591
Serial Correlation test:			
r_1	− ·0070	·1354	·0151
T-ratio	− ·04	·86	·10
r_2	·1271	·1649	·1397
T-ratio	·80	1·04	·88
r_3	− ·0602	·0453	− ·0456
T-ratio	− ·38	·29	− ·29
χ^2 test			
χ^2	37·37	37·15	37·11
Degrees of freedom	38	38	38
$p(\chi^2)$	·4983	·5085	·5102
Specimen values of q_x for integral ages, simulated standard errors, and standard errors as percentage of value of q_x:			
Age 20	·000337	·000221	·000351
(standard error)	·000085	·000043	·000062
percentage s.e.	25·08	19·48	17·71
Age 30	·000834	·000592	·000867

Table 15.2 (*cont.*)

Method Optimization criterion	L_1 Maximum likelihood	L_2 Normal approximation	L_3 $-\frac{1}{2}\chi^2$
standard error)	·000167	·000093	·000123
percentage s.e.	20·00	15·71	14·21
Age 40	·002058	·001582	·002142
(standard error)	·000312	·000191	·000231
percentage s.e.	15·14	12·04	10·80
Age 50	·005065	·004219	·005278
(standard error)	·000534	·000360	·000400
percentage s.e.	10·55	8·53	7·57
Age 60	·012385	·011174	·012915
(standard error)	·000797	·000602	·000619
percentage s.e.	6·43	5·39	4·79
Age 70	·029805	·029091	·031079
(standard error)	·001197	·001043	·001106
percentage s.e.	4·02	3·58	3·56
Age 80	·069119	·072531	·071953
(standard error)	·003698	·003295	·003418
percentage s.e.	5·35	4·54	4·75
Age 90	·148020	·163851	·153371
(standard error)	·011265	·009987	·009621
percentage s.e.	7·61	6·09	6·27
Age 100	·274188	·308174	·281679
(standard error)	·022362	·018529	·018047
percentage s.e.	8·16	6·01	6·41
Age 110	·416258	·456408	·423220
(standard error)	·027193	·019417	·020871
percentage s.e.	6·53	4·25	4·93

15.3 *Formulae of higher orders*

Since a simple two-parameter formula fits the data satisfactorily in this case, it is hardly necessary to try a graduation formula of a higher order. For completeness, however, we show in Tables 15.7 and 15.8 the values of the L_1 criterion and the values of the parameter estimates for higher order formulae, for μ_x, formula GM(r, s), maximum likelihood (i.e. corresponding to graduation Aa1), for the following formulae

Table 15.7 GM(0, 2), GM(0, 3), GM(1, 2)
Table 15.8 GM(0, 4), GM(1, 3), GM(2, 2)

The improvement in the value of the criterion as compared with the GM(0, 2) is at the most ·46 when one parameter is added, and 1·41 when two parameters are added. Such a small improvement does not justify the use of a formula with more parameters.

Further, it can be seen that the values of the added s-parameters, b_2 and b_3, are well within one standard error away from zero (T-ratios all less than 1·0),

Table 15.3. *Pensioners' widows, 1979-82.*
Statistics for graduations Ba.
Function: q_x (initial), formula: GM(0, 2)

Method Optimization criterion	L_1 Maximum likelihood	L_2 Normal approximation	L_3 $-\frac{1}{2}\chi^2$
Values of criteria at optimum point:			
L_1	− 3003·81	− 3004·65	− 3004·25
L_2	156·76	157·91	155·16
L_3	− 30·14	− 32·34	− 29·68
Values of parameters at optimum point:			
b_0	− 3·530580	− 3·565433	− 3·495853
(standard error)	·038071	·037083	·035817
T-ratio	− 92·74	− 96·15	− 97·60
b_1	4·160519	4·354299	4·088510
(standard error)	·184697	·155070	·150720
T-ratio	22·53	28·08	27·13
Comparison of total actual deaths (A) and total expected (E):			
Total A-E	1·87	20·24	− 20·69
Ratio 100A/E	100·27	103·01	97·10
(Ages grouped so that each $E_x \geqslant 5$)			
Signs test:			
Number of +	19	21	17
Number of −	23	20	25
p(pos)	·3220	·6224	·1400
Runs test:			
Number of runs	21	17	21
p(runs)	·4599	·1025	·5330
Kolmogorov–Smirnov test:			
Max deviation	·0224	·0305	·0251
p(KS)	·9950	·9096	·9797
Serial Correlation test:			
r_1	− ·1190	− ·0256	− ·0497
T-ratio	− ·77	− ·16	− ·32
r_2	·1329	·0951	·1370
T-ratio	·86	·61	·89
r_3	− ·0036	− ·1210	− ·0073
T-ratio	− ·02	− ·78	− ·05
χ^2 test			
χ^2	39·85	37·37	37·11
Degrees of freedom	40	39	40
$p(\chi^2)$	·4769	·5442	·6009
Specimen values of q_x for integral ages, simulated standard errors, and standard errors as percentage of value of q_x:			
Age 20	·000457	·000363	·000508

Table 15.3 (*cont.*)

Method Optimization criterion	L_1 Maximum likelihood	L_2 Normal approximation	L_3 $-\frac{1}{2}\chi^2$
(standard error)	·000099	·000063	·000080
percentage s.e.	21·73	17·33	15·77
Age 30	·001050	·000868	·001152
(standard error)	·000184	·000122	·000147
percentage s.e.	17·50	14·08	12·76
Age 40	·002413	·002075	·002609
(standard error)	·000324	·000226	·000256
percentage s.e.	13·44	10·92	9·83
Age 50	·005545	·004956	·005909
(standard error)	·000531	·000391	·000417
percentage s.e.	9·58	7·89	7·05
Age 60	·012744	·011840	·013386
(standard error)	·000780	·000613	·000623
percentage s.e.	6·12	5·18	4·66
Age 70	·029288	·028285	·030323
(standard error)	·001162	·001002	·001060
percentage s.e.	3·97	3·54	3·49
Age 80	·067308	·067571	·068690
(standard error)	·003417	·002990	·003174
percentage s.e.	5·08	4·42	4·62
Age 90	·154684	·161425	·155603
(standard error)	·012697	·011240	·010926
percentage s.e.	8·21	6·96	7·02
Age 100	·355486	·385639	·352485
(standard error)	·042233	·038617	·034636
percentage s.e.	11·88	10·01	9·83
Age 110	·816961	·921278	·798480
(standard error)	·128758	·122375	·102193
percentage s.e.	15·76	13·28	12·80

and that the values of the added *r*-parameters, a_0 and a_1, are less than two standard errors away from zero. Therefore none of the added parameters is itself significantly different from zero.

These results show that in this case a simple two-parameter formula is the most complex that the data can support, and we see no reason to suggest any other. The choice between μ_x with a GM(0, 2) formula, and q_x with a LGM(0, 2) formula is discussed below.

15.4 μ_x, q_x *or* m_x ?

There is some choice as to whether a life table should be described in terms of a function for μ_x or one for q_x. The tables produced by the CMI Committee in the past have been defined in terms of q_x. Where values of μ_x have also been

Table 15.4. *Pensioners' widows, 1979-82. Statistics for graduations Bb. Function: q_x (initial), formula: LGM(0, 2)*

Method Optimization criterion	L_1 Maximum likelihood	L_2 Normal approximation	L_3 $-\frac{1}{2}\chi^2$
Values of criteria at optimum point:			
L_1	− 3003·00	− 3004·61	− 3003·46
L_2	159·66	161·59	158·20
L_3	− 30·04	− 32·96	− 29·56
Values of parameters at optimum point:			
b_0	− 3·488932	− 3·517671	− 3·451337
(standard error)	·039507	·038543	·037349
T-ratio	− 88·31	− 91·27	− 92·41
b_1	4·424580	4·788848	4·371442
(standard error)	·206191	·173164	·167053
T-ratio	21·46	27·66	26·17
Comparison of total actual deaths (A) and total expected (E):			
Total A-E	·00	9·70	− 24·10
Ratio 100A/E	100·00	101·42	96·64
(Ages grouped so that each $E_x \geqslant 5$)			
Signs test:			
Number of +	19	22	17
Number of −	21	18	23
p(pos)	·4373	·7852	·2148
Runs test:			
Number of runs	20	18	22
p(runs)	·4440	·2293	·7377
Kolmogorov–Smirnov test:			
Max deviation	·0242	·0481	·0256
p(KS)	·9873	·4035	·9748
Serial Correlation test:			
r_1	− ·0239	·1083	− ·0100
T-ratio	− ·15	·68	− ·06
r_2	·1159	·1487	·1291
T-ratio	·73	·94	·82
r_3	− ·0713	·0226	− ·0568
T-ratio	− ·45	·14	− ·36
χ^2 test			
χ^2	36·22	35·85	36·03
Degrees of freedom	38	38	38
$p(\chi^2)$	·5520	·5693	·5609
Specimen values of q_x for integral ages, simulated standard errors, and standard errors as percentage of value of q_x:			
Age 20	·000366	·000247	·000400
(standard error)	·000088	·000047	·000070
percentage s.e.	24·17	19·12	17·37

Table 15.4 (*cont.*)

Method Optimization criterion	L_1 Maximum likelihood	L_2 Normal approximation	L_3 $-\frac{1}{2}\chi^2$
Age 30	·000885	·000643	·000959
(standard error)	·000171	·000099	·000134
percentage s.e.	19·33	15·45	13·95
Age 40	·002142	·001674	·002296
(standard error)	·000315	·000199	·000244
percentage s.e.	14·70	11·88	10·64
Age 50	·005175	·004350	·005487
(standard error)	·000533	·000368	·000411
percentage s.e.	10·31	8·46	7·49
Age 60	·012446	·011257	·013053
(standard error)	·000793	·000606	·000623
percentage s.e.	6·37	5·38	4·77
Age 70	·029629	·028814	·030729
(standard error)	·001184	·001030	·001085
percentage s.e.	3·99	3·57	3·53
Age 80	·068880	·071764	·070630
(standard error)	·003634	·003269	·003374
percentage s.e.	5·28	4·55	4·78
Age 90	·151987	·167684	·154105
(standard error)	·012000	·010949	·010329
percentage s.e.	7·90	6·53	6·70
Age 100	·302761	·344261	·303968
(standard error)	·028338	·025211	·023320
percentage s.e.	9·36	7·32	7·67
Age 110	·512680	·577717	·511452
(standard error)	·044231	·035478	·035443
percentage s.e	8·63	6·14	6·93

given, they have been derived from an assumption about the local shape of l_x (see McCutcheon, 1983). The classical Gompertz and Makeham formulae, however, express μ_x in a simple way.

It should be noted that all the functions associated with a life table are wholly defined if the values of μ_x are known for all x, but that they are not wholly defined if only the values of q_x are known, even for all x. Two different survivorship functions may have the same values of q_x at every age x (not just at integer values) (see McCutcheon, 1971). Given μ_x for all x, we may calculate the value of q_x for any desired x, if necessary by approximate integration, as noted in § 14.2, or explicitly. Explicit integration is possible if μ_x is represented by any GM(r, 2) formula. When q_x is used as the basis of the life table, some further assumption is required if values of μ_x are to be calculated.

If the life table is defined in terms of a known function for m_x, at least above some fixed age x_0, then above this age the whole table is uniquely specified, and values of l_x and μ_x can be calculated, using approximate methods but to any

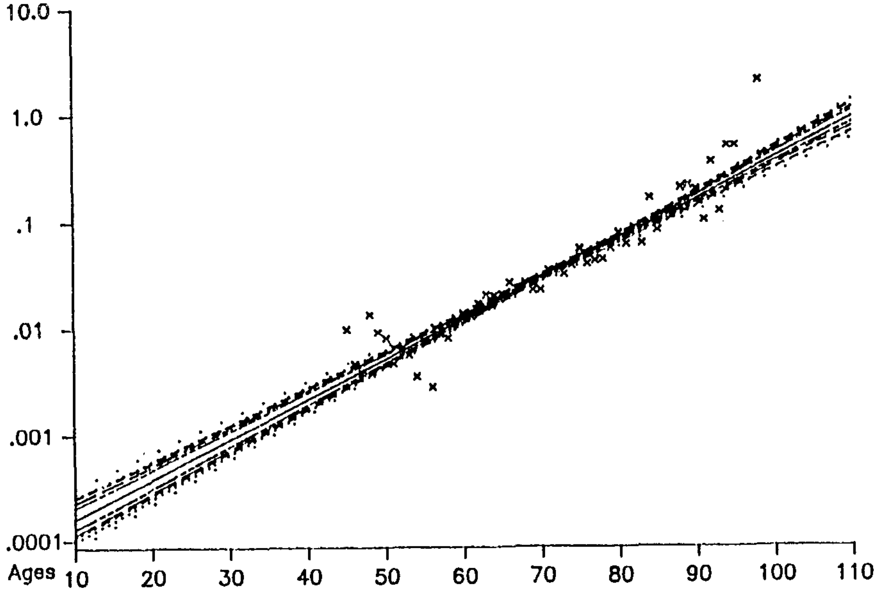

Figure 15.2. Widows of pensioners sheaf for μ_x = GM(0, 2).

Table 15.5. *Pensioners' widows, 1979-82.*
Details of graduation Aa1.
Function: μ_x formula: GM(0, 2), criterion: maximum likelihood.
Ages (nearest birthday) grouped so that each $E_x \geqslant 5$

Age	R_x	μ_x	A_x	E_x	Dev_x	$(V_x)^{1/2}$	Z_x	100A/*E*
17	·5	·00029499	0	·00	·00			
18	·0	·00032158	0	·00	·00			
19	·0	·00035058	0	·00	·00			
20	4·0	·00038219	0	·00	·00			
21	4·0	·00041665	0	·00	·00			
22	3·5	·00045422	0	·00	·00			
23	4·5	·00049518	0	·00	·00			
24	10·5	·00053983	0	·01	− ·01			
25	16·5	·00058850	0	·01	− ·01			
26	13·5	·00064156	0	·01	− ·01			
27	20·5	·00069941	0	·01	− ·01			
28	29·5	·00076248	0	·02	− ·02			
29	36·5	·00083123	0	·03	− ·03			
30	36·0	·00090618	0	·03	− ·03			
31	44·5	·00098789	0	·04	− ·04			

Table 15.5 (*cont.*)

Age	R_x	μ_x	A_x	E_x	Dev_x	$(V_x)^{1/2}$	Z_x	100A/*E*
32	50·0	·00107696	0	·05	− ·05			
33	64·0	·00117407	0	·08	− ·08			
34	73·0	·00127993	0	·09	− ·09			
35	79·5	·00139534	0	·11	− ·11			
36	80·0	·00152115	0	·12	− ·12			
37	93·5	·00165831	0	·16	− ·16			
38	106·5	·00180784	0	·19	− ·19			
39	122·0	·00197085	0	·24	− ·24			
40	115·5	·00214855	0	·25	− ·25			
41	127·0	·00234229	0	·30	− ·30			
42	157·0	·00255348	0	·40	− ·40			
43	184·5	·00278373	0	·51	− ·51			
44	191·0	·00303473	0	·58	− ·58			
45	206·5	·00330836	2	·68	1·32			
46	219·5	·00360667	1	·79	·21			
47	265·5	·00393188	1	1·04	− ·04			
17–47	2359·0		4	5·78	− 1·78	2·40	− ·74	69·2
48	301·5	·00428640	4	1·29	2·71			
49	330·5	·00467290	3	1·54	1·46			
50	378·5	·00509424	3	1·93	1·07			
51	437·5	·00555358	2	2·43	− ·43			
48–51	1448·0		12	7·19	4·81	2·68	1·79	166·8
52	480·0	·00605433	3	2·91	·09			
53	541·5	·00660024	3	3·57	− ·57			
52–53	1021·5		6	6·48	− ·48	2·55	− ·19	92·6
54	576·0	·00719537	2	4·14	− 2.14			
55	671·0	·00784416	5	5·26	− ·26			
55–55	1247·0		7	9·41	− 2·41	3·07	− ·79	74·4
56	719·5	·00855145	2	6·15	− 4·15	2·48	− 1·67	32·5
57	813·0	·00932251	7	7·58	− ·58	2·75	− ·21	92·4
58	879·0	·01016310	7	8·93	− 1·93	2·99	− ·65	78·4
59	934·0	·01107949	10	10·35	− ·35	3·22	− ·11	96·6
60	1029·0	·01207850	14	12·43	1·57	3·53	·45	112·6
61	1091·0	·01316759	14	14·37	− ·37	3·79	− ·10	97·5
62	1074·5	·01435488	18	15·42	2·58	3·93	·66	116·7
63	995·5	·01564923	20	15·58	4·42	3·95	1·12	128·4
64	963·5	·01706028	19	16·44	2·56	4·05	·63	115·6
65	1029·0	·01859857	21	19·14	1·86	4·37	·43	109·7
66	1108·5	·02027556	29	22·48	6·52	4·74	1·38	129·0
67	1130·5	·02210376	26	24·99	1·01	5·00	·20	104·0
68	1146·5	·02409681	30	27·63	2·37	5·26	·45	108·6
69	1037·0	·02626956	23	27·24	− 4·24	5·22	− ·81	84·4
70	941·0	·02863823	21	26·95	− 5·95	5·19	− 1·15	77·9
71	908·5	·03122048	31	28·36	2·64	5·33	·50	109·3

Table 15.5 (*cont.*)

Age	R_x	μ_x	A_x	E_x	Dev_x	$(V_x)^{1/2}$	Z_x	100A/E
72	844·5	·03403555	29	28·74	·26	5·36	·05	100·9
73	766·0	·03710446	24	28·42	− 4·42	5·33	− ·83	84·4
74	682·0	·04045009	26	27·59	− 1·59	5·25	− ·30	94·2
75	607·0	·04409738	33	26·77	6·23	5·17	1·20	123·3
76	533·0	·04807354	21	25·62	− 4·62	5·06	− ·91	82·0
77	500·5	·05240823	21	26·23	− 5·23	5·12	− 1·02	80·1
78	462·5	·05713376	20	26·42	− 6·42	5·14	− 1·25	75·7
79	382·5	·06228538	21	23·82	− 2·82	4·88	− ·58	88·1
80	323·5	·06790151	25	21·97	3·03	4·69	·65	113·8
81	282·0	·07402404	17	20·87	− 3·87	4·57	− ·85	81·4
82	243·5	·08069862	21	19·65	1·35	4·43	·30	106·9
83	213·5	·08797503	13	18·78	− 5·78	4·33	− 1·33	69·2
84	171·0	·09590754	28	16·40	11·60	4·05	2·86	170·7
85	132·5	·10455530	11	13·85	− 2·85	3·72	− ·77	79·4
86	99·5	·11398282	11	11·34	− ·34	3·37	− ·10	97·0
87	77·5	·12426039	10	9·63	·37	3·10	·12	103·8
88	59·0	·13546467	12	7·99	4·01	2·83	1·42	150·1
89	42·0	·14767921	9	6·20	2·80	2·49	1·12	145·1
90	30·5	·16099511	6	4·91	1·09			
91	19·5	·17551167	2	3·42	− 1·42			
90–91	50·0		8	8·33	− ·33	2·89	− ·13	96·0
92	8·5	·19133716	3	1·63	1·37			
93	8·0	·20858960	1	1·67	− ·67			
94	8·0	·22739766	4	1·82	2·18			
92–94	24·5		8	5·11	2·89	2·26	1·28	156·4
95	4·0	·24790159	2	·99	1·01			
96	2·5	·27025431	0	·68	− ·68			
97	2·5	·29462253	0	·74	− ·74			
98	·5	·32118798	1	·16	·84			
99	·5	·35014877	0	·18	− ·18			
100	1·0	·38172090	0	·38	− ·38			
101	·5	·41613981	0	·21	− ·21			
102	·0	·45366219	0	·00	·00			
103	1·0	·49456788	0	·49	− ·49			
104	·0	·53916195	0	·00	·00			
105	·0	·58777696	0	·00	·00			
106	·0	·64077548	0	·00	·00			
107	·0	·69855276	0	·00	·00			
108	2·0	·76153968	0	1·52	− 1·52			
95–108	14·5		3	5·35	− 2·35	2·31	− 1·01	56·1
Tot.	28386.5		692	692.00	·00			100·0

$\chi^2 = \Sigma z_x^2 = 38.29$

Table 15.6. *Pensioners' widows, 1979-82.*
Details of graduation Bb1.
Function: q_x (initial), formula: LGM(0, 2),
criterion: maximum likelihood.
Ages (nearest birthday) grouped so that each $E_x \geqslant 5$

Age	R_x	$q_{x-1/2}$	A_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
17	·5	·00026827	0	·00	·00			
18	·0	·00029309	0	·00	·00			
19	·0	·00032020	0	·00	·00			
20	4·0	·00034981	0	·00	·00			
21	4·0	·00038217	0	·00	·00			
22	3·5	·00041751	0	·00	·00			
23	4·5	·00045613	0	·00	·00			
24	10·5	·00049831	0	·01	− ·01			
25	16·5	·00054439	0	·01	− ·01			
26	13·5	·00059473	0	·01	− ·01			
27	20·5	·00064972	0	·01	− ·01			
28	29·5	·00070979	0	·02	− ·02			
29	36·5	·00077542	0	·03	− ·03			
30	36·0	·00084710	0	·03	− ·03			
31	44·5	·00092541	0	·04	− ·04			
32	50·0	·00101095	0	·05	− ·05			
33	64·0	·00110438	0	·07	− ·07			
34	73·0	·00120644	0	·09	− ·09			
35	79·5	·00131792	0	·10	− ·10			
36	80·0	·00143968	0	·12	− ·12			
37	93·5	·00157268	0	·15	− ·15			
38	106·5	·00171794	0	·18	− ·18			
39	122·0	·00187660	0	·23	− ·23			
40	115·5	·00204987	0	·24	− ·24			
41	127·0	·00223911	0	·28	− ·28			
42	157·0	·00244578	0	·38	− ·38			
43	184·5	·00267147	0	·49	− ·49			
44	191·0	·00291793	0	·56	− ·56			
45	207·5	·00318705	2	·66	1·34			
46	220·0	·00348091	1	·77	·23			
47	266·0	·00380175	1	1·01	− ·01			
17–47	2361·0		4	5·54	− 1·54	2·35	− ·66	72·1
48	303·5	·00415205	4	1·26	2·74			
49	332·0	·00453448	3	1·51	1·49			
50	380·0	·00495195	3	1·88	1·12			
51	438·5	·00540765	2	2·37	− ·37			
48–51	1454·0		12	7·02	4·98	2·64	1·88	171·0
52	481·5	·00590504	3	2·84	·16			
53	543·0	·00644788	3	3·50	− ·50			
52–53	1024·5		6	6·34	− ·34	2·51	− ·14	94·6

Table 15.6 (*cont.*)

Age	R_x	$q_{x-1/2}$	A_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
54	577·0	·00704027	2	4·06	− 2·06			
55	673·5	·00768667	5	5·18	− ·18			
54–55	1250·5		7	9·24	− 2·24	3·03	− ·74	75·8
56	720·5	·00839191	2	6·05	− 4·05	2·45	− 1·65	33·1
57	816·5	·00916126	7	7·48	− ·48	2·72	− ·18	93·6
58	882·5	·01000043	7	8·83	− 1·83	2·96	− ·62	79·3
59	939·0	·01091562	10	10·25	− ·25	3·18	− ·08	97·6
60	1036·0	·01191355	14	12·34	1·66	3·49	·47	113·4
61	1098·0	·01300152	14	14·28	− ·28	3·75	− ·07	98·1
62	1083·5	·01418742	18	15·37	2·63	3·90	·68	117·1
63	1005·5	·01547980	20	15·56	4·44	3·91	1·13	128·5
64	973·0	·01688788	19	16·43	2·57	4·02	·64	115·6
65	1039·5	·01842165	21	19·15	1·85	4·34	·43	109·7
66	1123·0	·02009186	29	22·56	6·44	4·70	1·37	128·5
67	1143·5	·02191013	26	25·05	·95	4·95	·19	103·8
68	1161·5	·02388894	30	27·75	2·25	5·20	·43	108·1
69	1048·5	·02604171	23	27·30	− 4·30	5·16	− ·83	84·2
70	951·5	·02838283	21	27·01	− 6·01	5·12	− 1·17	77·8
71	924·0	·03092773	31	28·58	2·42	5·26	·46	108·5
72	859·0	·03369291	29	28·94	·06	5·29	·01	100·2
73	778·0	·03669595	24	28·55	− 4·55	5·24	− ·87	84·1
74	695·0	·03995559	26	27·77	− 1·77	5·16	− ·34	93·6
75	623·5	·04349169	33	27·12	5·88	5·10	1·16	121·7
76	543·5	·04732533	21	25·72	− 4·72	4·95	− ·95	81·6
77	511·0	·05147869	21	26·31	− 5·31	5·00	− 1·06	79·8
78	472·5	·05597515	20	26·45	− 6·45	5·00	− 1·29	75·6
79	393·0	·06083916	21	23·91	− 2·91	4·74	− ·61	87·8
80	336·0	·06609624	25	22·21	2·79	4·55	·61	112·6
81	290·5	·07177287	17	20·85	− 3·85	4·40	− ·88	81·5
82	254·0	·07789637	21	19·79	1·21	4·27	·28	106·1
83	220·0	·08449475	13	18·59	− 5·59	4·13	− 1·35	69·9
84	185·0	·09159654	28	16·95	11·05	3·92	2·82	165·2
85	138·0	·09923054	11	13·69	− 2·69	3·51	− ·77	80·3
86	105·0	·10742555	11	11·28	− ·28	3·17	− ·09	97·5
87	82·5	·11621003	10	9·59	·41	2·91	1·14	104·3
88	65·0	·12561175	12	8·16	3·84	2·67	1·44	147·0
89	46·5	·13565735	9	6·31	2·69	2·34	1·15	142·7
90	33·5	·14637184	6	4·90	1·10			
91	20·5	·15777810	2	3·23	− 1·23			
90–91	54·0		8	8·14	− ·14	2·63	− ·05	98·3
92	10·0	·16989631	3	1·70	1·30			
93	8·5	·18274332	1	1·55	− ·55			
94	10·0	·19633201	4	1·96	2·04			

Table 15.6 (*cont.*)

Age	R_x	$q_{x-1/2}$	A_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
95	5·0	·21067068	2	1·05	·95			
96	2·5	·22576237	0	·56	− ·56			
97	2·5	·24160425	0	·60	− ·60			
98	1·0	·25818707	1	·26	·74			
99	·5	·27549462	0	·14	− ·14			
100	1·0	·29350334	0	·29	− ·29			
101	·5	·31218201	0	·16	− ·16			
102	·0	·33149165	0	·00	·00			
103	1·0	·35138549	0	·35	− ·35			
104	·0	·37180921	0	·00	·00			
105	·0	·39270130	0	·00	·00			
106	·0	·41399368	0	·00	·00			
107	·0	·43561243	0	·00	·00			
108	2·0	·45747879	0	·91	− ·91			
92–108	44·5		11	9·55	1·45	2·70	·54	115·2
Tot.	28732·50		692	692·00	·00			100·0

$\chi^2 = \Sigma z_x^2 = 36{\cdot}22$

desired degree of accuracy. There are, however, theoretical reasons for preferring μ_x to m_x as the basis of a graduation.

For most practical actuarial calculations a table of values of q_x for integer values of x is needed. How such a table of values has been calculated may be of little practical importance, and many different tables may be in use that have been constructed in different ways. In some circumstances, however, a simple formula for q_x may be desirable, so that values of q_x can be calculated directly by a computer function or subroutine. In this case, approximate integration of μ_x on each occasion that a value of q_x is required may be inconveniently slow. This is less of a problem if μ_x can be integrated explicitly to give an expression for q_x.

In this paper we have discussed the fitting of a mathematical formula to mortality data for a particular experience. In order to produce a mortality table for practical use, it may be necessary to extend the range of ages over which the formula has been fitted, either upwards to the highest ages, or downwards to lower ages, or both. It is convenient if the formula itself provides a reasonable function for higher ages, and many of the formulae we have used do this. For a downwards extension a variety of methods may be used.

For the lower ages of the A1967-70 table the CMI Committee used an *ad hoc* extension based on a table of values of q_x at integral ages only. For the lower ages for the annuitants table *a*eg 1967-70, on which the *a*(90) tables were based, the Committee used for males the A1967-70 formula rates, with a blending function; the values of q_x for all x could therefore be derived by means of a formula, although rather a complicated one. For females the rates for lower ages were derived from the A1967-70 rates ratioed by the rates for ELT12 Females and ELT12 Males, both of which were based on formulae, so again a rather

Table 15.7. *Pensioners' widows, 1979-82.*
Statistics for graduations Aa1.
Function: μ_x, formula: GM(r, s).
Optimisation criterion: Maximum likelihood (L_l)

Number of parameters	2	3	3
r	0	0	1
s	2	3	2
Values of criterion at optimum point:			
L_l	− 3003·23	− 3003·21	− 3002·79
Values of parameters at optimum point:			
$a_0 \times 100$	–	–	− ·132331
(standard error)	–	–	·085059
T-ratio	–	–	− 1.56
b_0	− 3·553013	− 3·618036	− 3·489439
(standard error)	·039234	·310230	·056184
T-ratio	− 90·56	− 11·66	− 62·11
b_1	4·316579	4·325999	4·07591
(standard error)	·196615	·202828	·262517
T-ratio	21·95	21·33	15·52
b_2	–	− ·070109	–
(standard error)	–	·331634	–
T-ratio	–	− ·21	–

complicated formula was required. The pensioners' tables, PA(90), for males were not extended below age 50, and for females the formula provided reasonable values down to age 20. In these cases a simple formula for q_x sufficed.

Where a table has had to be extended by an elaborate method the advantage of a simple formula for q_x may be lost, and it may be more convenient to use a table of values of q_x for integral ages. In such cases it may be preferable to use the theoretically nicer method of defining μ_x by formula, which generally may be of the GM(r, s) type. In cases, however, where a simple formula for q_x both fits the data and provides reasonable values for the whole range of ages desired, then the practical advantages may outweigh the theoretical niceties; in general a LGM(r, s) formula may be preferable.

16. EXAMPLE 2—MALE PENSIONERS

16.1 *The Data*

For our second example we use a much larger experience, that for male life office pensioners retiring at or after their normal retirement age. For this graduation we ignore the select data, subdivided by duration since retirement, and use only the aggregate data, as was done when the experience for 1967-70

Table 15.8. *Pensioners' widows, 1979-82.*
Statistics for graduations Bb1.
Funciton: μ_x, formula: GM(r, s).
Optimisation criterion: Maximum likelihood (L_l)

Number of parameters	4	4	4
r	0	1	2
s	4	3	2
Values of criterion at optimum point:			
L_1	− 3003·19	− 3002·43	− 3001·82
Values of parameters at optimum point:			
$a_0 \times 100$	–	− ·421281	·855473
(standard error)	–	·808794	·524312
T-ratio	–	− ·52	1·63
$a_1 \times 100$	–	–	1·491302
(standard error)	–	–	·819679
T-ratio	–	–	1·82
b_0	− 3·628966	− 2·926007	− 3·919935
(standard error)	·318213	1·075632	·295883
T-ratio	− 11·40	− 2·72	− 13·25
b_1	4·492413	3·623105	5·094109
(standard error)	·992057	1·064041	·775866
T-ratio	4·53	3·40	6·57
b_2	− ·082429	·482083	–
(standard error)	·340834	·846736	–
T-ratio	− ·24	·57	–
b_3	·066381	–	–
(standard error)	·386927	–	–
T-ratio	·17	–	–

for the same investigation was used to provide the basis for the Peg1967-70 graduated tables described by the C.M.I. Committee (1976).

On that occasion the Committee graduated q_x using a LGM(0,2) formula. It was of interest to see whether the same formula would be satisfactory for the 1979-82 data.

The total central exposed to risk for this experience was 1,377,059·5 years (initial 1,419,772·5), and there were 85,426 deaths, more than 100 times the number of deaths in our first example. The numbers at each age appear in Tables 16.5 (central) and 16·6 (initial). The extreme limits of age observed in the exposed to risk were 19 to 108 (nearest birthday), but it must be suspected that the tiny amount of data below age 30 is the result of errors, and even from age 30 to about 50 is as likely to be erroneous as valid. Outside the range of ages from 54 to 97 inclusive the central exposed to risk at each age was less than 100, and outside the range of ages from 60 to 100 inclusive the number of deaths at each age was less than 10, so the bulk of the deaths were concentrated into the same sort of span of years as the pensioners' widows in the first example.

As for the pensioners' widows there was no reason to suppose that a large number of duplicates existed in this experience, and no information from which variance ratios could be derived, so it was assumed that each observation represented one 'life'.

We graduated this experience both by using 'initial' exposures and graduating q_x, and by using 'central' exposures and graduating μ_x, as in the first example.

The crude rates of μ_x or q_x for each age were calculated, and those for μ_x are plotted in Figure 16.1. Confidence intervals for μ_x and q_x were also calculated, using the methods described in § 2.6, and using the normal approximation for ages where the number of deaths exceeded 60 (ages 64 to 96 inclusive) and the exact Poisson or Binomial method for other ages. The upper and lower limits of the 95% confidence intervals for μ_x are also shown in Figure 16.1 along with the graph of the GM(1,3) formula for μ_x discussed below.

It can be seen that the confidence intervals are extremely narrow at ages from 65 to 92 or so. There can be little doubt about where the graduated rates should lie. Below 65 there are some irregularities (age 64 shows a rather high crude rate), and below age 55 the confidence intervals become very wide. Above age 92 the confidence intervals widen somewhat, and the crude rates appear to fall

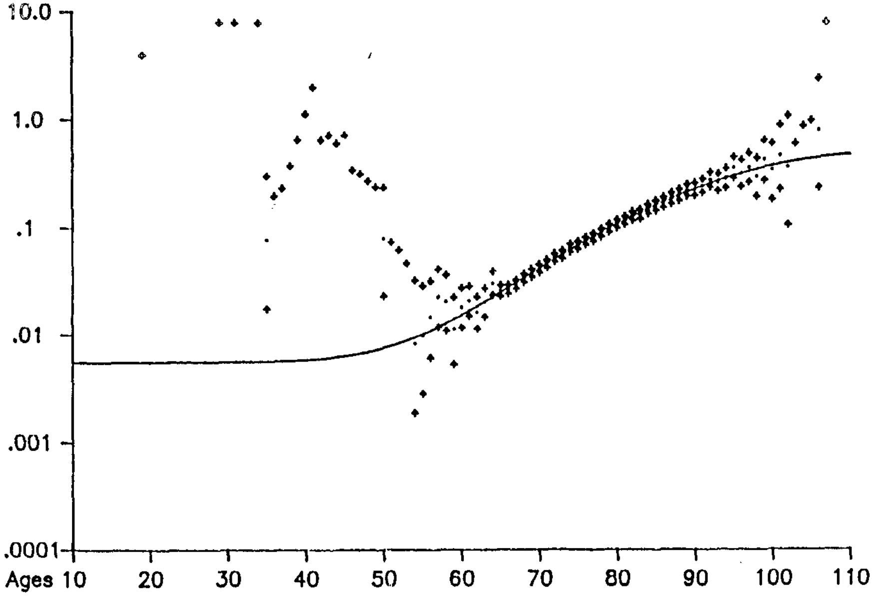

Figure 16.1. Male pensioners (N/L, lives) crude rates and gates: μ_x = GM(1,3).

below a linear projection of the rates at younger ages. This suggests that a GM(0,2) formula would not fit the data well at these ages.

16.2 *Choice of Order of Formula*

We first explored which order of formula might be suitable. We began this exploration with graduations of μ_x, using a GM(r, s) formula and a maximum likelihood (L_1) criterion. We tried values of r and s with $s \geqslant 2$, $r + s \leqslant 6$, a total of 15 formulae. The values of the L_1 criterion (plus a large constant) are shown below.

r \ s	2	3	4	5	6
0	−155·9	−58·5	−55·4	−53·4	−53·4
1	−98·7	−52·6	−51·5	−46·9	
2	−53·3	−50·9	−50·9		
3	−54·0	−50·7			
4	−52·2				

It can be seen that the value of the criterion either increases or remains the same as one moves down a column, or along a row, in each case as one extra parameter is fitted. Along a NE/SW diagonal the number of parameters is constant. It is therefore reasonable to prefer that entry along such a diagonal that shows the highest value. The Akaike criterion (Edwards' level of support) suggests that one step down or to the right is not worth making unless the value of L_1 is improved by 2. It is easily seen that the entry for GM(1,3), a value of −52·6, is the first possible stopping point, since that value is much more than 2 greater than its neighbours above and to the left, is greater than its neighbours to north-east and south-west, and differs by less than 2 from its neighbours below and to the right.

The other possible contender is GM(1,5), which shows a value of −46·9. This is 5·7 higher than the value for GM(1,3), but the formula requires two extra parameters. If twice the difference between the values of L_1, i.e. 11·4, is compared with $\chi^2(2)$, it is readily seen to be very significant. Thus a GM(1,5) formula could be preferred to a GM(1,3) on this criterion.

There are, however, other considerations to take into account. Table 16.1 summarises these considerations for each of the 15 formulae investigated. The first two columns of this table show the order (r, s) of the formula and the value of L_1 already discussed. In the next column, headed 'Good fit', is an indication of whether the graduation satisfies the following tests: signs test, runs test, serial correlation test, and Kolmogorov–Smirnov test (with a value of p(KS) exceeding ·9). It can be seen that the formulae of too low an order, GM(0, 2) and GM(1,2), do not pass these tests.

Next is shown the values of χ^2 and of $p(\chi^2)$. A graduation is satisfactory if the value of $p(\chi^2)$ is greater than say ·05. It can be seen that the first four formulae in the table do not satisfy this test, and that GM(0,6) is marginal.

If the estimated value of the highest order parameter in either the r-poly-

Table 16.1. *Comparison of GM(r, s) formulae for graduating μ_x*

Order (r, s)	Value of L_1	Good fit	χ^2	$p(\chi^2)$	Signf. pars.	Good shape	Tight sheaf
0,2	−155·9	No	243·8	·00	Yes	Yes	Yes
0,3	−58·5	Yes	65·1	·02	Yes	Yes	Yes
1,2	−98·7	No	130·9	·00	Yes	No	No
0,4	−55·4	Yes	60·0	·04	Yes	Yes	No
1,3	−52·6	Yes	54·7	·11	Yes	Yes	Yes
2,2	−53·3	Yes	54·1	·19	Yes	No	No
0,5	−53·4	Yes	56·4	·08	Yes	No	No
1,4	−51·5	Yes	52·0	·19	Yes	Yes	Yes
2,3	−50·9	Yes	51·4	·21	No	No	Yes
3,2	−54·0	Yes	54·9	·15	No	No	No
0,6	−53·4	Yes	56·7	·05	No	No	No
1,5	−46·9	Yes	43·3	·46	Yes	No	Yes
2,4	−50·9	Yes	51·1	·19	No	No	No
3,3	−50·7	Yes	50·9	·19	No	No	No
4,2	−52·2	Yes	51·9	·17	No	No	No

nomial or the s-polynomial is not significantly different from zero, this indicates possible over-parameterization. It is therefore appropriate to note whether these parameters are significantly different from zero. This is indicated in the next column, headed 'Signf. pars.'. It can be seen that when the number of parameters does not exceed 4 the highest order ones remain significant, but that when the number does exceed 4 one or other or both of the highest order parameters are found to be non-significant for all formulae except GM(0,5), GM(2,3) and GM(1,5).

A formula which provides a 'good' shape for the ages beyond where the bulk of the data is found may be preferable to one that extrapolates to values that are clearly unreasonable. A good shape is defined for this purpose as one where the values of μ_x continue to rise as age increases up to age 110, and neither rise unreasonably nor fall away to zero as age reduces down to 20. The column headed 'Good shape' indicates whether or not this criterion is satisfied. It can be seen that most of the higher-order formulae do not provide a good enough shape.

Finally, a graduation may be preferred if the 'sheaf' of quantile plots of simulated values of μ_x, described in § 11.3, is fairly tight, indicating that the relative standard errors of the graduated estimates of μ_x are reasonably small. The tightness of the sheaf is associated with the size of the standard errors of the parameter estimates. What is tight for this purpose is a relative matter, but in this case there is a clear division between tight and loose sheaves. It can be seen that many of the high-order formulae and even some of the low-order formulae do not have tight sheaves.

There is only one formula that satisfies all these criteria, and that is the GM(1,3) formula already identified as potentially the best through its high value for L_1. The GM(1,5) formula has many good features, but it has a rather poor shape, in fact reaching a maximum value of μ_x at $x = 98$, and falling sharply thereafter.

The graph of the GM(1,3) formula for μ_x is shown in Figure 16.1, and the sheaf in Figure 16.2, in which the quantiles plotted are the same as those described for the pensioners' widows, viz: numbers 1, 3, 5, 10, 20, 81, 91, 96, 98 and 100 out of 100 simulations. It can be seen how much tighter is the sheaf for this large experience than is that for pensioners' widows, shown in Figure 15.2. The standard errors of the estimates of q_x shown in Table 16.3 are also much smaller than those in Table 15.1.

As a contrast, in Figure 16.3 we show the sheaf for a GM(3,3) formula. The maximum likelihood estimates of the parameters for this formula and their standard errors and T-ratios are

	Parameter Estimate	Standard Error	T-ratio
$a_0 \times 100$	−23·632738	43·996884	−·5371
$a_1 \times 100$	−34·505888	58·025867	−·5947
$a_2 \times 100$	−10·747659	16·846927	−·6380
b_0	−2·341549	2·215710	−1·0568
b_1	3·286758	1·861814	1·7654
b_2	−·571860	·621756	−·9198

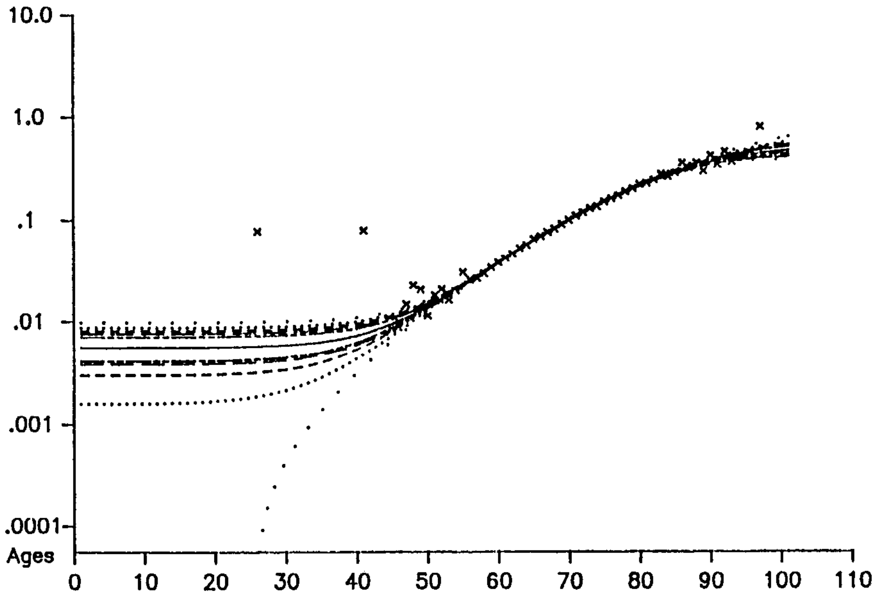

Figure 16.2. Male pensioners (N/L, lives) sheaf for μ_x = GM(1,3).

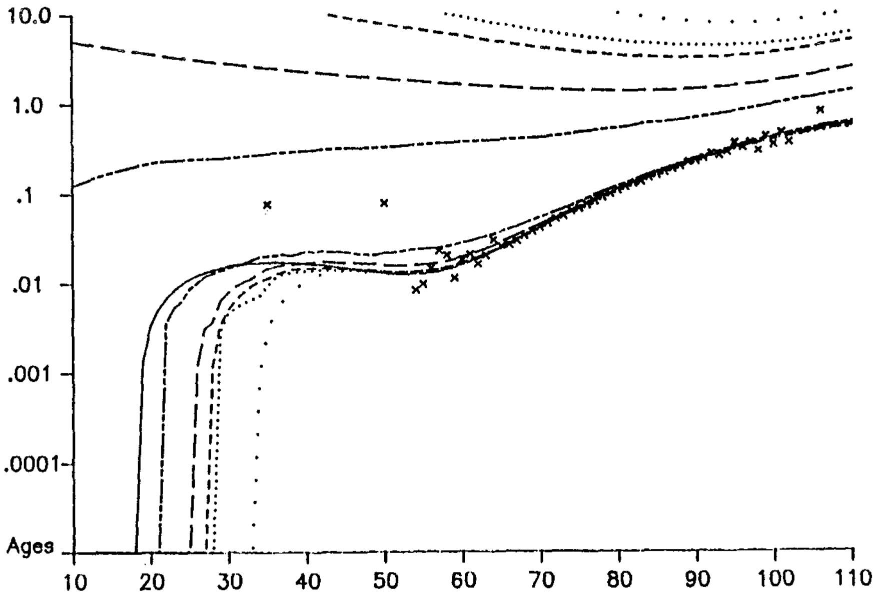

Figure 16.3. Male pensioners (N/L, lives) sheaf for μ_x = GM(3,3).

The effect of over-parameterization can readily be appreciated.

A similar exploration was carried out for graduations of q_x with a LGM(r,s) formula. Values of L_1 (plus a large constant) are

s	2	3	4	5	6
r					
0	−77·7	−23·0	−20·7	−18·5	−18·5
1	−68·7	−18·0	−16·4	−12·9	
2	−16·9	−16·5	−16·0		
3	−16·6	−16·0			
4	−15·7				

These values cannot be compared directly with the values of L_1 for the formulae graduating μ_x shown above, since a constant adjustment needs to be allowed for. The tables show a similar pattern. The LGM(1,3) formula is potentially good, but in this case the LGM(2,2) formula would appear to be better. LGM(1,5) is again a contender.

Table 16.2 shows how these formulae match up to the same criteria as discussed above for the graduations of μ_x. The entries are similar to those in Table 16.1, but they are not identical. Although formula LGM(2,2) has a high value of L_1, it has neither a good shape nor a tight sheaf, so we do not prefer it to LGM(1,3).

Table 16.2. *Comparison of LGM(r, s) formulae for graduating q_x*

Order (r, s)	Value of L_1	Good fit	χ^2	$p(\chi^2)$	Signf. pars.	Good shape	Tight sheaf
0,2	−77·7	No	169·5	·00	Yes	Yes	Yes
0,3	−23·0	Yes	64·6	·02	Yes	Yes	Yes
1,2	−68·7	No	148·7	·00	No	No	No
0,4	−20·7	Yes	60·0	·04	No	Yes	No
1,3	−18·0	Yes	55·4	·10	Yes	Yes	Yes
2,2	−16·9	Yes	52·3	·18	Yes	No	No
0,5	−18·5	Yes	56·7	·06	Yes	No	No
1,4	−16·4	Yes	52·6	·18	No	Yes	Yes
2,3	−16·5	Yes	51·8	·17	Yes	No	No
3,2	−16·6	Yes	52·3	·16	No	No	Yes
0,6	−18·5	Yes	56·9	·05	No	No	No
1,5	−12·9	Yes	43·6	·45	Yes	No	Yes
2,4	−16.0	Yes	51·7	·15	No	No	No
3,3	−16·0	Yes	52·0	·14	No	No	No
4,2	−15·7	Yes	50·7	·16	No	No	No

16.3 *Alternative Criteria*

For the pensioners' widows we compared different methods of fitting each of the functions. We do this again, but using only the combinations

(Aa) μ_x using Gompertz–Makeham GM(1,3) and
(Bb) q_x using logistic LGM(1,3)

In each case we calculated the values described in § 15. The results of each of the graduations are summarized in Tables 16.3 and 16.4, which correspond to Tables 15.1 and 15.4, showing

Table 16.3	Graduation Aa μ_x	GM(1,3)
Table 16.4	Graduation Bb q_x (initial)	LGM(1,3)

We can observe that all six graduations shown are very much closer together than was the case in the first example. The parameters and the values of μ_x or q_x (at least for the main age range of about 60 to 100) for the normal approximation (L_2) are quite close to those for the exact maximum likelihood (L_1) criterion. Those for the minimum $\frac{1}{2}\chi^2$ (L_3) criterion are a little further away, but not substantially so. All the graduations fit the data satisfactorily according to all the relevant tests.

The standard errors for the parameter estimates appear quite wide, but in fact they are closely correlated, in such a way that a variation in one parameter requires a corresponding variation in another so that the values of μ_x or q_x produced are not so very different. This is seen from the very low standard errors for the estimates of μ_x or q_x at the ages where the data is densest. For example,

Table 16.3. *Male pensioners, normal or late, 1979-82. Statistics for graduations Aa. Function: μ_x, formula: GM(1,3)*

Method	L_1	L_2	L_3
Optimization criterion	Maximum likelihood	Normal approximation	$-\frac{1}{2}\chi^2$
Values of criteria at optimum point:			
L_1	−309752·58	−309752·65	−309754·12
L_2	77·27	77·39	75·98
L_3	−45·70	−44·59	−42·89
Values of parameters at optimum point:			
$a_0 \times 100$	·557291	·617109	·831983
(standard error)	·183966	·120454	·141503
T-ratio	3·03	5·12	5·88
b_0	−4·993529	−5·068128	−5·344232
(standard error)	·265676	·195440	·237788
T-ratio	−18·80	−25·93	−22·47
b_1	5·882482	5·963083	6·275090
(standard error)	·273044	·192761	·240608
T-ratio	21·54	30·94	26·08
b_2	−1·668855	−1·726347	−1·937322
(standard error)	·215576	·163880	·196673
T-ratio	−7·74	−10·53	−9·85
Comparison of total actual deaths (A) and total expected (E):			
Total A–E	1·00	−5·11	−42·01
Ratio $100A/E$	100·00	99·99	99·95
(Ages grouped so that each $E_x \geqslant 5$)			
Signs test:			
Number of +	23	24	25
Number of −	24	23	23
p(pos)	·5000	·6146	·6673
Runs test:			
Number of runs	29	27	25
p(runs)	·9304	·8119	·5616
Kolmogorov–Smirnov test:			
Max deviation	·0019	·0019	·0026
P(KS)	·9984	·9980	·9284
Serial Correlation test:			
r_1	·0018	·0035	·0702
T-ratio	·01	·02	·49
r_2	−·1140	−·1135	−·0778
T-ratio	−·78	−·78	−·54
r_3	−·0611	−·0660	−·0728
T-ratio	−·42	−·45	−·50
χ^2 test			
χ^2	54·72	54·73	57·71
Degrees of freedom	43	43	44
$p(\chi^2)$	·1085	·1083	·0805

Table 16.3 (*cont.*)

Method	L_1	L_2	L_3
Optimization criterion	Maximum likelihood	Normal approximation	$-\frac{1}{2}\chi^2$
Specimen values of q_x for integral ages, simulated standard errors, and standard errors as percentage of value of q_x:			
Age 20	·005561	·006155	·008287
(standard error)	·001626	·001006	·001438
percentage s.e.	29·24	16·35	17·35
Age 30	·005600	·006189	·008306
(standard error)	·001613	·000995	·001429
percentage s.e.	28·79	16·08	17·20
Age 40	·005906	·006466	·008496
(standard error)	·001526	·000937	·001371
percentage s.e.	25·83	14·49	16·14
Age 50	·007729	·008184	·009864
(standard error)	·001216	·000736	·001134
percentage s.e.	15·73	9·00	11·50
Age 60	·015886	·016105	·016929
(standard error)	·000588	·000356	·000579
percentage s.e.	3·70	2·21	3·42
Age 70	·042799	·042781	·042677
(standard error)	·000225	·000202	·000225
percentage ss.e.	·53	·47	·53
Age 80	·106334	·106369	·106575
(standard error)	·000557	·000513	·000547
percentage s.e.	·52	·48	·51
Age 90	·209121	·208980	·209213
(standard error)	·002342	·002151	·002380
percentage s.e.	1·12	1·03	1·14
Age 100	·317159	·314823	·308279
(standard error)	·011132	·010908	·013032
percentage s.e.	3·51	3·46	4·23
Age 110	·379986	·372433	·348706
(standard error)	·030153	·028303	·034134
percentage s.e.	7·94	7·60	9·79

at ages 70 to 80 the values of μ_x and q_x can be estimated with a standard error of only one half per cent of their value.

In this example we have shown that the simplest formulae that adequately fit the data are a GM(1,3) for μ_x and a LGM(1,3) for q_x. Although certain higher order formulae fit the data just as well, and in terms of the log-likelihood criterion rather better, the other considerations that should be taken into account make them less satisfactory than the GM(1,3) and LGM(1,3) formulae.

Table 16.4. *Male pensioners, normal or late, 1979-82. Statistics for graduations Bb. Function: q_x (initial), formula: LGM(1,3)*

Method	L_1	L_2	L_3
Optimization criterion	Maximum likelihood	Normal approximation	$-\frac{1}{2}\chi^2$
Values of criteria at optimum point:			
L_1	−309717·99	−309718·05	−309719·56
L_2	81·31	81·41	79·82
L_3	−49·24	−48·22	−46·29
Values of parameters at optimum point:			
$a_0 \times 100$	·538616	·591097	·833373
(standard error)	·195921	·095333	·148560
T-ratio	2·75	6·20	5·61
b_0	−4·700716	−4·758472	−5·085900
(standard error)	·282191	·168455	·252404
T-ratio	−16·66	−28·25	−20·15
b_1	5·897192	5·961896	6·312090
(standard error)	·281004	·153143	·246157
T-ratio	20·99	38·93	25·64
b_2	−1·464466	−1·508395	−1·767744
(standard error)	·233190	·147569	·212793
T-ratio	−6·28	−10·22	8·31
Comparison of total actual deaths (A) and total expected (E):			
Total A–E	−·02	−7·80	−36·36
Ratio $100A/E$	100·00	99·99	99·96
(Ages grouped so that each $E_x \geqslant 5$)			
Signs test:			
Number of +	24	23	25
Number of −	23	24	23
p(pos)	·6146	·5000	·6673
Runs test:			
Number of runs	29	29	25
p(runs)	·9304	·9304	·5616
Kolmogorov–Smirnov test:			
Max deviation	·0018	·0019	·0026
p(KS)	·9989	·9984	·9420
Serial Correlation test:			
r_1	·0029	·0042	·0680
T-ratio	·02	·03	·47
r_2	−·1085	−·1091	−·0752
T-ratio	−·74	−·75	−·52
r_3	−·0626	−·0680	−·0740
T-ratio	−·43	−·47	−·51
χ^2			
χ^2 test	55·40	55·33	58·20
Degrees of freedom	43	43	44
$p(\chi^2)$	·0973	·0985	·0741

Table 16.4 (*cont.*)

Method Optimization criterion	L_1 Maximum likelihood	L_2 Normal approximation	L_3 $-\frac{1}{2}\chi^2$
Specimen values of q_x for integral ages, simulated standard errors, and standard errors as percentage of value of q_x:			
Age 20	·005363	·005881	·008267
(standard error)	·001704	·000791	·001496
percentage s.e.	31·78	13·45	18·10
Age 30	·005411	·005923	·008289
(standard error)	·001686	·000780	·001485
percentage s.e.	31·16	13·17	17·92
Age 40	·005751	·006238	·008491
(standard error)	·001597	·000727	·001420
percentage s.e.	27·77	11·66	16·72
Age 50	·007653	·008050	·009882
(standard error)	·001255	·000564	·001165
percentage s.e.	16·40	7·01	11·79
Age 60	·015880	·016078	·016939
(standard error)	·000595	·000295	·000587
percentage s.e.	3·75	1·83	3·47
Age 70	·042785	·042771	·042653
(standard error)	·000229	·000197	·000225
percentage s.e.	·54	·46	·53
Age 80	·106434	·106454	·106706
(standard error)	·000562	·000498	·000548
percentage s.e.	·53	·47	·51
Age 90	·209901	·209904	·209644
(standard error)	·002338	·002161	·002432
percentage s.e.	1·11	1·03	1·16
Age 100	·322789	·321598	·313193
(standard error)	·010429	·009613	·012394
percentage s.e.	3·23	2·99	3·96
Age 110	·404906	·400729	·373526
(standard error)	·027012	·023393	·031796
percentage s.e.	6·67	5·84	8·51

Tables 16.5 and 16.6 show detailed results for these two graduations. In each table the value of z_x exceeds 2.0 on three occasions, for age group 56–57, and ages 64 and 74. No value of z_x is less than -2.0. An unusually high number of deaths at or about age 64 in this investigation in previous periods has been noticed by the CMI Committee. The exposed to risk changes sharply between ages 64 and 65, multiplying over tenfold, and it is possible that the census method used to calculate the exposed to risk does not give a good enough estimate in these circumstances.

Table 16.5. *Male pensioners, normal or late, 1979-82*
Details of graduation Aa1.
Function: μ_x, formula: GM(1,3), criterion: maximum likelihood.
Ages grouped so that each $E_x \geqslant 5$

Age	R_x	μ_x	A_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
19	1·0	·00557567	0	·01	−·01			
20–28	no data							
29	·5	·00560356	0	·00	·00			
30	no data							
31	·5	·00562094	0	·00	·00			
32–33	no data							
34	·5	·00566522	0	·00	·00			
35	13·0	·00568707	1	·07	·93			
36	20·0	·00571371	0	·11	−·11			
37	17·0	·00574611	0	·10	−·10			
38	10·5	·00578540	0	·06	−·06			
39	6·0	·00583291	0	·03	−·03			
40	3·5	·00589018	0	·02	−·02			
41	2·0	·00595904	0	·01	−·01			
42	6·0	·00604160	0	·04	−·04			
43	5·5	·00614028	0	·03	−·03			
44	6·5	·00625791	0	·04	−·04			
45	5·5	·00639773	0	·04	−·04			
46	11·5	·00656343	0	·08	−·08			
47	12·5	·00675925	0	·08	−·08			
48	14·5	·00699000	0	·10	−·10			
49	16·5	·00726111	0	·12	−·12			
50	25·5	·00757873	2	·19	1·81			
51	53·5	·00794975	0	·43	−43			
52	63·5	·00838188	0	·53	−·53			
53	84·0	·00888373	0	·75	−·75			
54	121·0	·00946484	1	1·15	−·15			
55	206·0	·01013573	2	2·09	−·09			
19–55	706·5		6	6·09	−·09	2·47	−·03	98·6
56	341·0	·01090802	5	3·72	1·28			
57	442·5	·01179438	10	5·22	4·78			
56–57	783·5		15	8·94	6·06	2·99	2·03	167·8
58	537·5	·01280865	11	6·88	4·12	2·62	1·57	159·8
59	617·0	·01396584	7	8·62	−1·62	2·94	−·55	81·2
60	1380·2	·01528213	25	21·09	3·91	4·59	·85	118·5
61	2459·5	·01677490	51	41·26	9·74	6·42	1·52	123·6
62	2649·0	·01846273	43	48·91	−5·91	6·99	−·84	87·9
63	2884·8	·02036532	58	58·75	−·75	7·66	−·10	98·7
64	3271·8	·02250346	99	73·63	25·37	8·58	2·96	134·5

Table 16.5 (*cont.*)

Age	R_x	μ_x	A_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$.
65	36460·2	·02489899	937	907·82	29·18	30·13	·97	103·2
66	90619·0	·02757463	2408	2498·79	−90·79	49·99	−1·82	96·4
67	101939·0	·03055392	3008	3114·64	−106·64	55·81	−1·91	96·6
68	105445·2	·03386100	3556	3570·48	−14·48	59·75	−·24	99·6
69	104575·8	·03752047	3945	3923·73	21·27	62·64	·34	100·5
70	101021.8	·04155713	4209	4198·17	10·83	64·79	·17	100·3
71	96954·0	·04599576	4448	4459·47	−11·47	66·78	−·17	99·7
72	92197·5	·05086080	4806	4689·24	116·76	68·48	1·71	102·5
73	86210·8	·05617606	4808	4842·98	−34·98	69·59	−·50	99·3
74	80050·2	·06196438	5149	4960·26	188·74	70·43	2·68	103·8
75	73819·2	·06824723	5047	5037·96	9·04	70·98	·13	100·2
76	67097·2	·07504433	5037	5035·27	1·73	70·96	·02	100·0
77	60212·0	·08237325	4867	4959·86	−92·86	70·43	−1·32	98·1
78	52777·0	·09024892	4727	4763·07	−36·07	69·02	−·52	99·2
79	45130·2	·09868328	4456	4453·60	2·40	66·74	·04	100·1
80	37312·0	·10768474	4049	4017·93	31·07	63·39	·49	100·8
81	29974·2	·11725781	3509	3514·71	−5·71	59·28	−·10	99·8
82	23539·0	·12740261	3016	2998·93	17·07	54·76	·31	100·6
83	18308·5	·13811453	2448	2528·67	−80·67	50·29	−1·60	96·8
84	14281·0	·14938379	2126	2133·35	−7·35	46·19	−·16	99·7
85	11134·0	·16119512	1775	1794·75	−19·75	42·36	−·47	98·9
86	8578·5	·17352745	1467	1488·61	−21·61	38·58	−·56	98·5
87	6622·2	·18635370	1234	1234·08	−·08	35·13	−·00	100·0
88	5104·8	·19964058	1021	1019·12	1·88	31·92	·06	100·2
89	3827·8	·21334846	842	816·64	25·36	28·58	·89	103·1
90	2787·5	·22743141	627	633·97	−6·97	25·18	−·28	98·9
91	1989·8	·24183718	482	481·20	·80	21·94	·04	100·2
92	1323·2	·25650742	365	339·42	25·58	18·42	1·39	107·5
93	895·0	·27137788	233	242·88	−9·88	15·58	−·63	95·9
94	579·5	·28637877	165	165·96	−·96	12·88	−·07	99·4
95	376·8	·30143518	134	113·57	20·43	10·66	1·92	118·0
96	240·8	·31646765	76	76·19	−·19	8·73	−·02	99·8
97	159·0	·33139275	57	52·69	4·31	7·26	·59	108·2
98	95·0	·34612381	28	32·88	−4·88	5·73	−·85	85·2
99	59·0	·36057173	25	21·27	3·73	4·61	·81	117·5
100	32·2	·37464579	11	12·08	−1·08	3·48	−·31	91·0
101	17·2	·38825464	8	6·70	1·30	2·59	·50	119·4
102	5·5	·40130719	2	2·21	−·21			
103	6·5	·41371369	0	2·69	2·69			
104	4·5	·42538666	0	1·91	−1·91			
105	4·0	·43624199	0	1·74	−1·74			
106	2·5	·44619989	2	1·12	·88			
107	·5	·45518589	0	·23	−·23			
108	·0	·46313176	1	·00	1·00			
102–108	23·5		5	9·90	−4·90	3·15	−1·56	50·5
T.	1377059·5		85426	85425·00	1·00			100·0

$\chi^2 = \Sigma z_x^2 = 54{\cdot}72$

Table 16.6. *Male pensioners, normal or late, 1979-82*
Details of graduation Bb1.
Function: q_x (initial), formula: LGM(1,3),
criterion: maximum likelihood.
Ages grouped so that each $E_x \geqslant 5$

Age	R_x	$q_{x-1/2}$	A_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
19	1·0	·00536131	0	·01	−·01			
20–28	no data							
29	·5	·00539601	0	·00	·00			
30	no data							
31	·5	·00541655	0	·00	·00			
32–33	no data							
34	·5	·00546754	0	·00	·00			
35	13·5	·00549225	1	·07	·93			
36	20·0	·00552212	0	·11	−·11			
37	17·0	·00555812	0	·09	−·09			
38	10·5	·00560142	0	·06	−·06			
39	6·0	·00565334	0	·03	−·03			
40	3·5	·00571548	0	·02	−·02			
41	2·0	·00578963	0	·01	−·01			
42	6·0	·00587791	0	·04	−·04			
43	5·5	·00598273	0	·03	−·03			
44	6·5	·00610689	0	·04	−·04			
45	5·5	·00625357	0	·03	−·03			
46	11·5	·00642641	0	·07	−·07			
47	12·5	·00662957	0	·08	−·08			
48	14·5	·00686774	0	·10	−·10			
49	16·5	·00714622	0	·12	−·12			
50	26·5	·00747099	2	·20	1·80			
51	53·5	·00784874	0	·42	−·42			
52	63·5	·00828695	0	·53	−·53			
53	84·0	·00879393	0	·74	−·74			
54	121·5	·00937889	1	1·14	−·14			
55	207·0	·01005199	2	2·08	−·08			
19–55	709·5		6	6·04	−·04	2·45	−·01	99·4
56	343·5	·01082437	5	3·72	1·28			
57	447·5	·01170820	10	5·24	4·76			
56–57	791·0		15	8·96	6·04	2·98	2·03	167·5
58	543·0	·01271671	11	6·91	4·09	2·61	1·57	159·3
59	620·5	·01386421	7	8·60	−1·60	2·91	−·55	81·4
60	1392·8	·01516606	25	21·12	3·88	4·56	·85	118·4
61	2485·0	·01663869	51	41·35	9·65	6·38	1·51	123·3
62	2670·5	·01829954	43	48·87	−5·87	6·93	−·85	88·0
63	2913·8	·02016698	58	58·76	−·76	7·59	−·10	98·7
64	3321·2	·02226025	99	73·93	25·07	8·50	2·95	133·9
65	36928·8	·02459931	937	908·42	28·58	29·77	·96	103·1
66	91823·0	·02720468	2408	2498·02	−90·02	49·30	−1·83	96·4

Table 16.6 (*cont.*)

Age	R_x	$q_{x-1/2}$	A_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
67	103443·0	·03009725	3008	3113·35	−105·35	54·95	−1·92	96·6
68	107223·2	·03329807	3556	3570·33	−14·33	58·75	−·24	99·6
69	106548·2	·03682806	3945	3923·97	21·03	61·48	·34	100·5
70	103126·2	·04070770	4209	4198·03	10·97	63·46	·17	100·3
71	99178·0	·04495672	4448	4458·72	−10·72	65·26	−·16	99·8
72	94600·5	·04959370	4806	4691·59	114·41	66·78	1·71	102·4
73	88614·8	·05463572	4808	4841·53	−33·53	67·65	−·50	99·3
74	82624·8	·06009792	·5149	4965·58	183·42	68·32	2·68	103·7
75	76342·8	·06599309	5047	5038·09	8·91	68·60	·13	100·2
76	69615·8	·07233129	5037	5035·40	1·60	68·35	·02	100·0
77	62645·5	·07911940	4867	4956·47	−89·47	67·56	−1·32	98·2
78	55140·5	·08636078	4727	4761·98	−34·98	65·96	−·53	99·3
79	47358·2	·09405498	4456	4454·28	1·72	63·52	·03	100·0
80	39336·5	·10219737	4049	4020·09	28·91	60·08	·48	100·7
81	31728·8	·11077902	3509	3514·88	−5·88	55·91	−·11	99·8
82	25047·0	·11978654	3016	3000·29	15·71	51·39	·31	100·5
83	19532·5	·12920199	2448	2523·64	−75·64	46·88	−1·61	97·0
84	15344·0	·13900299	2126	2132·86	−6·86	42·85	−·16	99·7
85	12021·5	·14916281	1775	1793·16	−18·16	39·06	−·46	99·0
86	9312·0	·15965060	1467	1486·67	−19·67	35·35	−·56	98·7
87	7239·2	·17043174	1234	1233·80	·20	31·99	·01	100·0
88	5615·2	·18146817	1021	1018·99	2·01	28·88	·07	100·2
89	4248·8	·19271892	842	818·81	23·19	25·71	·90	102·8
90	3101·0	·20414059	627	633·04	−6·04	22·45	−·27	99·0
91	2230·8	·21568791	482	481·15	·85	19·43	·04	100·2
92	1505·8	·22731431	365	342·28	22·72	16·26	1·40	106·6
93	1011·5	·23897253	233	241·72	−8·72	13·56	−·64	96·4
94	662·0	·25061515	165	165·91	−·91	11·15	−·08	99·5
95	443·8	·26219519	134	116·35	17·65	9·27	1·91	115·2
96	278·8	·27366654	76	76·28	−·28	7·44	−·04	99·6
97	187·5	·28498448	57	53·43	3·57	6·18	·58	106·7
98	109·0	·29610607	28	32·28	−4·28	4·77	−·90	86·8
99	71·5	·30699042	25	21·95	3·05	3·90	·78	113·9
100	37·8	·31759902	11	11·99	−·99	2·86	−·35	91·7
101	21·2	·32789591	8	6·97	1·03	2·16	·48	114·8
102	6·5	·33784782	2	2·20	−·20			
103	6·5	·34742421	0	2·26	−2·26			
104	4·5	·35659733	0	1·60	−1·60			
105	4·0	·36534216	0	1·46	−1·46			
106	3·5	·37363631	2	1·31	·69			
107	·5	·38145996	0	·19	−·19			
108	·5	·38879569	1	·19	·81			
102–108	26·0		5	9·21	−4·21	2·44	−1·73	54·3
T.	1419772·5		85426	85426·02	−·02			100·0

$\chi^2 = \Sigma z_x^2 = 55{\cdot}40$

Table 16.7. *Male pensioners, normal or late, 1979-82.*
Lives and Amounts
Statistics for graduations Aa1.
Function: μ_x, criterion: maximum likelihood
(Amounts values in units of £324.416)

Date Formula	Lives GM(1,3)	Amounts GM(0,3)	Amounts GM(1,3)
Values of criterion at optimum point:			
L_1	−309752·58	−241553·12	−241552·04
Values of parameters at optimum point:			
$a_0 \times 100$	·557291	–	·200562
(standard error)	·183966	–	·125554
T-ratio	3·03	–	1·60
b_0	−4·993529	−4·379382	−4·716405
(standard error)	·265676	·092408	·239617
T-ratio	−18·80	−47·39	−19·68
b_1	5·882482	5·487972	5·832101
(standard error)	·273044	·057148	·232910
T-ratio	21·54	96·03	25·04
b_2	−1·668855	−1·005111	−1·277685
(standard error)	·215576	·092497	·202092
T-ratio	−7·74	−10·87	−6·32
Comparison of total actual deaths (*A*) and total expected (*E*):			
Total *A–E*	1·00	·57	·54
Ratio 100*A*/*E*	100·00	100·00	100·00
(Ages grouped so that each $E_x \geqslant 5$)			
Signs test:			
Number of +	23	24	24
Number of −	24	24	24
p(pos)	·5000	·5573	·5573
Runs test:			
Number of runs	29	27	29
p(runs)	·9304	·7660	·9052
Kolmogorov–Smirnov test:			
Max deviation	·0019	·0045	·0053
p(KS)	·9984	·5621	·3557
Serial Correlation test:			
r_1	·0018	−·0083	−·0518
T-ratio	·01	·06	−·36
r_2	−·1140	−·0295	−·0013
T-ratio	−·78	−·02	−·01
r_3	−·0611	−·3061	−·3198
T-ratio	−·42	−2·12	−2·22
χ^2 test			
χ^2	54·72	217·55	214·36
Degrees of freedom	43	45	44
$p(\chi^2)$	·1085	·0000	·0000

Table 16.7 (*cont.*)

Date Formula	Lives GM(1,3)	Amounts GM(0,3)	Amounts GM(1,3)
Specimen values of q_x for integral ages, simulated standard errors, and standard errors as percentage of value of q_x:			
Age 20	·005561	·000021	·002012
(standard error)	·001626	·000004	·001031
percentage s.e.	29·24	21·51	51·26
Age 30	·005600	·000128	·002069
(standard error)	·001613	·000018	·001018
percentage s.e.	28·79	14·25	49·19
Age 40	·005906	·000668	·002426
(standard error)	·001526	·000058	·000975
percentage s.e.	25·83	8·62	40·17
Age 50	·007729	·002964	·004237
(standard error)	·001216	·000131	·000765
percentage s.e.	15·73	4·43	18·05
Age 60	·015886	·011169	·011611
(standard error)	·000588	·000182	·000347
percentage s.e.	3·70	1·63	2·99
Age 70	·042799	·035536	·035405
(standard error)	·000225	·000155	·000191
percentage s.e.	·53	·44	·54
Age 80	·106334	·094480	·094850
(standard error)	·000557	·000519	·000626
percentage s.e.	·52	·55	·66
Age 90	·209121	·206884	·205732
(standard error)	·002342	·002857	·002936
percentage s.e.	1·12	1·38	1·43
Age 100	·317159	·369299	·354160
(standard error)	·011132	·011957	·013838
percentage s.e.	3·51	3·24	3·91
Age 110	·379986	·541793	·492450
(standard error)	·030153	·028839	·037474
percentage s.e.	7·94	5·32	7·61

16.4 *'Amounts' Data*

We have so far considered the data based on 'Lives', but 'Amounts' are also collected for this investigation, and the standard PA(90) tables are based on the graduation of the 1967-70 Amounts data. The existence of multiple pound amounts on one life is equivalent to there being multiple policies on one life, so it is an example of the presence of duplicates. We have no knowledge about the distribution of amounts per life other than the average amounts for each life at each age among the exposed to risk and the deaths.

Table 16.8. *Male pensioners, normal or late, 1979–82. Amounts values in units of £324.416 Details of graduation Aa1. Function: μ_x, formula: GM(1,3), maximum likelihood. Ages grouped so that each $E_x \geqslant 5$*

Age	R_x	μ_x	A_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
19	·6	·00201148	·00	·00	·00			
20	·3	·00201293	·00	·00	·00			
21–28	no data							
29	·4	·00205386	·00	·00	·00			
30	no data							
31	·3	·00207736	·00	·00	·00			
32–33	no data							
34	1·2	·00213374	·00	·00	·00			
35	33·4	·00216042	·03	·07	−·04			
36	62·7	·00219229	·00	·14	−·14			
37	44·6	·00223025	·00	·10	−·10			
38	32·4	·00227538	·00	·07	−·07			
39	4·2	·00232892	·00	·01	−·01			
40	2·4	·00239229	·00	·01	−·01			
41	0·7	·00246714	·00	·00	·00			
42	5·1	·00255536	·00	·01	−·01			
43	7·7	·00265909	·00	·02	−·02			
44	5·5	·00278082	·00	·02	−·02			
45	10·2	·00292334	·00	·03	−·03			
46	52·5	·00308984	·00	·16	−·16			
47	100·7	·00328394	·00	·33	−·33			
48	22·7	·00350971	·00	·08	−·08			
49	20·6	·00377174	·00	·08	−·08			
50	44·5	·00407517	·93	·18	·75			
51	105·6	·00442579	·00	·47	−·47			
52	167·5	·00483003	·00	·81	−·81			
53	172·4	·00529506	·00	·91	−·91			
54	228·0	·00582883	1·76	1·33	·43			
55	533·4	·00644014	2·48	3·44	−·96			
19–55	1659·7		5·20	8·27	−3·07	2·88	−1·07	62·9
56	834·1	·00713869	15·56	5·95	9·60	2·44	3·94	261·3
57	998·3	·00793515	21·61	7·92	13·69	2·81	4·86	272·8
58	1327·7	·00884120	13·64	11·74	1·90	3·43	·55	116·2
59	1567·1	·00986960	12·82	15·47	−2·65	3·93	−·67	82·9
60	5927·9	·01103424	27·54	65·41	−37·87	8·09	−4·68	42·1
61	11596·5	·01235020	168·79	143·22	25·57	11·97	2·14	117·9
62	11918·4	·01383376	163·18	164·88	−1·69	12·84	−·13	99·0
63	12431·6	·01550246	226·77	192·72	34·05	13·88	2·45	117·7
64	13226·2	·01737513	205·24	229·81	−24·57	15·16	−1·62	89·3
65	62014·0	·01947188	1228·81	1207·53	21·28	34·75	·61	101·8
66	132344·5	·02181415	2847·69	2886·98	−39·29	53·73	−·73	98·6
67	133769·6	·02442464	3140·18	3267·27	−127·10	57·16	−2·22	96·1

Table 16.8 (*cont.*)

Age	R_x	μ_x	A_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
68	126595·5	·02732733	3442·87	3459·52	−16·65	58·82	−·28	99·5
69	114842·0	·03054745	3491·83	3508·13	−16·30	59·23	−·28	99·5
70	103696·8	·03411137	3642·05	3537·24	104·81	59·47	1·76	103·0
71	94832·7	·03804654	3511·88	3608·06	−96·17	60·07	−1·60	97·3
72	82919·9	·04238142	3776·04	3514·26	261·78	59·28	4·42	107·4
73	72703·8	·04714531	3393·81	3427·64	−33·84	58·55	−·58	99·0
74	64109·3	·05236822	3561·40	3357·29	204·11	57·94	3·52	106·1
75	57064·8	·05808070	3363·00	3314·36	48·64	57·57	·84	101·5
76	51027·4	·06431362	3023·98	3281·76	−257·78	57·29	−4·50	92·1
77	44402·4	·07109797	3018·78	3156·92	−75·14	56·19	−1·34	97·6
78	37494·6	·07846456	2884·94	2942·00	−57·05	54·24	−1·05	98·1
79	31022·7	·08644377	2708·86	2681·72	27·14	51·79	·52	101·0
80	25081·6	·09506526	2350·73	2384·39	−33·66	48·83	−·69	98·6
81	19695·1	·10435758	2089·39	2055·33	34·06	45·34	·75	101·7
82	15051·9	·11434788	1826·29	1721·15	105·14	41·49	2·53	106·1
83	11611·7	·12506148	1352·02	1452·17	−100·15	38·11	−2·63	93·1
84	8989·5	·13652153	1159·77	1227·26	−67·49	35·03	−1·93	94·5
85	6855·0	·14874854	1037·46	1019·67	17·79	31·93	·56	101·7
86	5178·8	·16176002	837·76	837·72	·04	28·94	·00	100·0
87	3978·7	·17557003	805·02	698·54	106·48	26·43	4·03	115·2
88	3004·4	·19018875	550·85	571·40	−20·55	23·90	−·86	96·4
89	2287·2	·20562208	470·73	470·30	·42	21·69	·02	100·1
90	1649·1	·22187121	375·15	365·90	9·26	19·13	·48	102·5
91	1167·6	·23893221	284·56	278·97	5·59	16·70	·33	102·0
92	769·1	·25679570	207·36	197·50	9·87	14·05	·70	105·0
93	517·5	·27544648	124·39	142·54	−18·15	11·94	−1·52	87·3
94	336·9	·29486320	73·51	99·34	−25·83	9·97	−2·59	74·0
95	215·6	·31501814	59·64	67·93	−8·29	8·24	−1·01	87·8
96	121·1	·33587694	56·83	40·67	16·15	6·38	2·53	139·7
97	81·6	·35739848	40·52	29·16	11·36	5·40	2·10	139·0
98	58·8	·37953474	16·66	22·30	−5·64	4·72	−1·19	74·7
99	34·2	·40223078	20·91	13·77	7·14	3·71	1·92	151·8
100	17·1	·42542476	3·52	7·28	−3·76	2·70	−1·39	48·3
101	11·1	·44904807	5·50	4·97	·53			
102	6·0	·47302547	5·23	2·82	2·41			
101–102	17·0		10·73	7·79	2·95	2·79	1·06	137·8
103	6·6	·49727541	·00	3·30	−3·30			
104	2·3	·52171031	·00	1·20	−1·20			
105	1·5	·54623703	·00	·82	−·82			
106	1·4	·57075737	·28	·82	−·54			
107	0·4	·59516863	·00	·24	−·24			
108	·0	·61936429	·55	·00	·55			
103–108	12·3		·83	6·39	−5·56	2·53	−2·20	12·9
T.	1377059·5		61714·09	61713·55	0·54			100.0

$\chi^2 = \Sigma z_x^2 = 214{\cdot}36$

The CMI Committee, when graduating the corresponding 1967-70 experience, divided the exposed to risk and deaths at each age by a constant factor, being the average pound amount among the exposed to risk over all ages. This simply scales down the amounts so that tests like the χ^2 test can be used. It makes no difference to the estimation of the maximum likelihood parameters, but it changes their standard errors very considerably, and consequently the standard errors of the values of μ_x or q_x. It does not affect the signs test nor the runs test.

If we had only Amounts data and we were wishing to derive rates appropriate to Lives, it might be appropriate to divide the exposed to risk and the deaths at each age by the average amount per life at each age, using these average amounts as if they were variance ratios. If each life at each age had the same number of pounds, the average amount would equal the variance ratio; but if this were so, the average amount for each death would be the same as the average amount per life among the exposed to risk, and this is not the case. The Amounts experiences for pensioners regularly show lower mortality than the corresponding Lives experiences, and the average amount per death is less than the average amount per life in the exposed to risk, at least in total, if not at each single age.

However, we have already derived mortality rates for Lives, and we now wish to derive mortality rates appropriate to Amounts. The technique used by the Committee previously seems the only reasonable one. The resulting parameter estimates are at least asymptotically unbiassed, even though many of the tests we have used for Lives are invalidated. We therefore divided each pound amount by £324.416, and measured in 'units' of this amount. The deaths and exposed to risk are

	£	Units
Deaths	20,021,034	61,714.1
Exposed to risk (initial)	456,750,562.5	1,407,916.6
Exposed to risk (central)	446,740,045.5	1,377,059.5

We fitted μ_x using a GM(r, s) formula to the Amounts data, and using an L_1 criterion. Values of L_1 (plus a large constant) for various formulae are

r \ s	2	3	4	5	6
0	−114·7	−53·1	−52·2	−52·2	−51·7
1	−73·3	−52·0	−52·0	−49·2	
2	−55·9	−52·0	−51·9		
3	−52·3	−51·4			
4	−51-5				

The previous criteria for assessing differences in the value of L_1 are not strictly valid in this case, and the T-ratios of the parameter estimates cannot be validly compared with a normal distribution. However, if we were to ignore this point, we would conclude that a GM(1,3) formula (as for the Lives data) provided a satisfactory fit, although the value of χ^2 (214·4) would be thought unacceptably high. No higher order formula reduces the value of χ^2 substantially. The a_0 parameter in the GM(1,3) formula would be thought not significantly different from zero ($100a_0$ = ·200562, T-ratio = 1·5974). A GM(0,3) formula would fit almost equally well, and the s-parameters would have similar values. One could therefore conclude that a GM(0,3) formula was the most suitable for this experience.

However, a GM(1,3) formula was found to be appropriate for the Lives data, and there is an advantage in being able to compare the parameters and the resulting values of q_x. Table 16.7 gives the statistics of the GM(0,3) and GM(1,3) graduations (Aa1 basis), and compares them with the statistics for the GM(1,3) graduation of Lives (repeated from Table 16.3). Table 16.8 gives details of the graduation for each age.

It is interesting to note that, although some of the tests we use are not strictly valid for Amounts data, they make a fairly good pretence of being suitable. The values of z_x at certain ages are relatively large, and the coincidence of large values of opposite sign at ages 57 and 60 contributes to the large negative value of r_3. The graduated values of q_x are clearly substantially lower for the Amounts data than for Lives data.

17. EXAMPLE 3—MALE ASSURED LIVES (UNITED KINGDOM — PERMANENT)

17.1 *Introduction*

For our third example we move to a more complicated experience, that for United Kingdom male lives assured for permanent (whole-life and endowment) assurances. This investigation includes data for a select period of five years and an ultimate period thereafter, so there are initially six different sets of data, for durations 0, 1, 2, 3, 4, and 5 and over. We shall abbreviate this last to 'duration 5+', and we shall use a similar notation when we group durations; for example durations 2, 3, 4 and 5+ combined will be referred to as 'duration 2+'.

The experience of this class of life for previous periods has formed the basis of several sets of standard tables for assured lives, A1924-29, A1949-52, A1967-70 and A1967-70(5). For the first of these a standard table with three years selection was constructed. For A1949-52 this was reduced to two years selection. The 1967-70 data provided the basis for two sets of tables, one with two years and the other with five years selection. The question of what select period might be appropriate therefore requires investigation.

17.2 *Duplicates and Variance Ratios*

Previous investigations by the CMI Committee had discovered that there was often a considerable number of duplicate policies in the data. That is, one life was insured under two or more policies. Some of the statistical methods and tests described above are less reliable when there are duplicates, and a method of allowing for duplicates has been described (§§ 3.2 and 6.2). This is to adjust the numbers of deaths and the exposed to risk by dividing each of them by the appropriate 'variance ratio', r_x, and then using the adjusted numbers in place of the original numbers. While this method is theoretically correct only for certain of the tests described above, it is the only practicable way of dealing with duplicates, given the data available to the CMI Committee.

If duplicate cases could be eliminated reliably from the original exposed to risk, so that each life was recorded only once in the investigation, this adjustment would not be required. This cannot be done. It is, however, possible to count the number of policies on each life among the deaths, since a copy of the death certificate for each life assured who has died is supplied to the Bureau for the Cause of Death investigation. By matching cases that correspond appropriately, a count of duplicates (not just within offices, but also across offices) can be obtained. The result of such an investigation for two of the four years of the quadrennium has been described by the CMI Committee (1986). We have available also the corresponding results for the full four years of the quadrennium.

To use variance ratios derived from records of the deaths as if they applied also to the exposed to risk is to make two assumptions: first that mortality rates for lives assured of the same age do not depend on the number of policies for which they are insured, so that the variance ratios derived from the deaths are unbiassed estimates of the variance ratios that would have been derived from the exposed to risk; secondly, that the experience is large enough for the standard deviations of the variance ratios derived from the deaths (as estimates of the variance ratios from the exposed to risk) to be small.

We have no evidence about the first assumption. It is not valid for pensioners and annuitants when 'duplicates' are interpreted as pound amounts on one life, since the 'Amounts' experience for these investigations is markedly different from that for 'lives', as we have seen in § 16.4 (though we do not strictly know whether the difference is statistically significant). As regards the second assumption, we know that the number of duplicates among deaths is large only for the 5+ experience (and hence the experiences for 0+, 1+, etc which include the 5+ data).

This is the large part of the experience, so it is reasonable to assume that the possible error caused by using the variance ratios derived from the deaths is small, at least for the main range of ages where the numbers are large.

Since one life may have policies in the data for more than one of the different durations, when the experiences for different durations are combined a separate count of the duplicates and separate calculation of the variance ratios is required; the ratios cannot be derived in any way by adding totals for two or more durations together. The count of duplicates, and hence variance ratios, has been done for each duration separately, and for duration 2+. The values of the ratios for each of the select durations are small (i.e. close to 1·0) and we have ignored duplicates in investigating these experiences. The variance ratios for duration 5+ are shown in Table 17.6 and those for duration 2+ are in the accompanying CMI Report.

17.3 *The Data*

The numbers of deaths and the exposed to risk (central and initial) for each of the separate durations are as shown below. For durations 2+ and 5+ we show both the unadjusted and adjusted numbers.

Duration	Deaths	Central exposed	Initial exposed
0	1,795	1,799,039·7	1,799,937·2
1	2,287	1,776,058·3	1,777,201·8
2	2,388	1,704,683·3	1,705,877·3
3	2,509	1,649,973·7	1,651,228·2
4	2,606	1,571,019·8	1,572,322·8
5+	83,438	17,313,471·2	17,355,190·2
2+	90,941	22,239,148·0	22,284,618·5
Adjusted by variance ratios			
5+	53,239·3	10,827,136·0	10,853,755·6
2+	56,539·1	13,440,029·9	13,468,299·4

We have noted previously three age ranges of interest: first the extreme limits of the data; secondly the continuous range over which the exposed to risk at each age exceeds 100; thirdly the continuous range over which the deaths at each age exceed 10. We note each range for each duration below.

Duration	Range of data	Central exposed ⩾ 100	Deaths ⩾ 10
0	10–100	10– 76	17– 67
1	10– 88	12– 76	18– 68
2	10– 89	12– 77	19– 69
3	10–100	13– 77	20– 72
4	10– 97	14– 76	21– 72
5+	10–108	15–100	22–100
2+	10–108	10–100	18–100

The same age ranges apply also to the adjusted data.

It can be seen there is adequate data in the select durations for a wide range of ages from below 20 to about 70, and for the ultimate data for the whole range of adult life. The data for individual ages for duration 0 is shown in Table 17.4, that for durations 2 to 4 combined in Table 17.5.

17.4 *Comparison of data for different durations*

In § 13 we discussed the tests that can be applied to the data for two different experiences in order to see whether they are significantly different, or whether they are sufficiently similar for the data to be pooled before graduating. We applied these tests to this experience, in order to distinguish or to amalgamate durations. We first compared each single duration with each other single duration, and then each group of one or more consecutive durations with each other neighbouring group of one or more consecutive durations. The results are summarized in Table 17.1.

In order to apply a χ^2 test, the data for duration 5+ has been adjusted by dividing the numbers of deaths and the exposed to risk at each age by the corresponding variance ratio. This does not affect the signs test nor the runs test. Central exposures have been used. For each comparison ages have been grouped so that there were at least 5 actual deaths in each age group in each experience.

Table 17.1 shows the results for each comparison in one line. The durations or groups of durations compared are identified as I and II. Thus the first line compares duration 0 with duration 1, and the last compares durations 0 to 4 combined with duration 5+.

The next two columns show the ratios of actual to expected deaths for the two experiences, where the expected deaths at each age are calculated from the pooled experience. It can be seen that in every case the earlier duration or group of durations shows lighter mortality in aggregate than the later.

The next two columns show the number of ages or age groups where the first duration showed a crude mortality rate higher (+) or lower (−) than the second. The column headed $p(+)$ shows the probability of this number or fewer + signs being observed if the probability of $a+$ were one half. It can be seen that in no case does the number of +s exceed the number of −s, i.e. the second experience always shows a majority of ages or age groups with higher mortality rates than the first. In most cases the value of $p(+)$ is very small, indicating significant difference in the level of mortality between the two durations. We shall comment on the exceptions below.

The next two columns show the number of runs of similar sign among the +s and −s of the previous test, and the probability, p(runs), of as small or smaller a number of runs. Although the number of runs is often small, this is because the number of +s is also small. In very few cases is the value of p(runs) less than ·05 and in no case less than ·01. This shows that the experiences of the different durations are in some sense reasonably 'parallel', and overlap to the extent expected, considering that their levels differ.

The last two columns show the values of χ^2 and of $p(\chi^2)$. The number of degrees of freedom is the sum of the numbers of + and − signs; it is at the lowest 56 and the highest 62, so the significant and non-significant values of χ^2 are readily identified. It can be seen that the value of $p(\chi^2)$ is in most cases very small, indicating a significant difference between the mortality rates of the two durations being compared.

Inspection of the table shows that duration 0 has significantly lower mortality than any other duration singly or any combination of durations, and that duration 5+ has significantly higher mortality than any other. Any group of durations including 5+ is significantly higher than any earlier group of durations.

Durations 1, 2, 3 and 4 are not so readily separated. Durations 1 and 3, and durations 3 and 4 are not significantly different (at a 5% level) either by the signs

Table 17.1. *Male assured lives, permanent, U.K. Comparison of data for different durations*

Durations I v II		I A/E	II A/E	Numbers +	−	$p(+)$	runs	p(runs)	χ^2	$p(\chi^2)$
0	1	88·9	110·9	11	47	·0000	13	·0165	102·3	·0003
0	2	87·2	112·4	14	44	·0001	19	·1679	141·1	·0000
0	3	85·2	114·2	6	51	·0000	10	·1470	146·6	·0000
0	4	83·3	116·0	3	53	·0000	7	1·0000	165·0	·0000
0	5+	68·1	101·6	1	55	·0000	3	1·0000	385·7	·0000
1	2	98·0	102·0	22	37	·0337	28	4847	43·0	·9417
1	3	96·0	104·0	23	34	·0924	30	·7142	67·7	·1565
1	4	93·8	106·1	15	41	·0003	24	·6804	70·0	·0984
1	5+	82·0	101·0	7	50	·0000	11	·1582	184·8	·0000
2	3	98·0	102·0	28	31	·3974	33	·7904	80·9	·0308
2	4	96·0	104·0	21	36	·0314	31	·8759	64·2	·2393
2	5+	85·1	100·8	12	46	·0000	17	·1635	149·9	·0000
3	4	98·2	101·8	26	32	·2559	28	·3753	54·1	·6224
3	5+	87·3	100·7	12	46	·0000	21	·7375	126·7	·0000
4	5+	90·0	100·6	16	41	·0006	19	·0719	113·1	·0000
0	1–2	84·8	107·4	16	43	·0003	19	·0593	144·3	·0000
1	2–3	96·1	101·9	19	40	·0043	26	·4575	47·0	·8698
2	3–4	96·1	102·0	25	34	·1488	35	·9380	76·8	·0596
3	4+	87·7	100·6	12	47	·0000	21	·7266	124·2	·0000
0	1–3	81·9	105·8	6	53	·0000	11	·4328	151·8	·0000
1	2–4	94·1	102·0	18	41	·0019	22	·1352	55·8	·5925
2	3+	85·6	100·7	11	48	·0000	19	·6135	144·7	·0000
0	1–4	79·6	104·9	3	56	·0000	7	1·0000	167·3	·0000
1	2+	83·4	100·8	10	50	·0000	13	·0364	161·6	·0000
0	1+	68·7	101·3	0	59	·0000	1	1·0000	348·0	·0000
0–1	2	95·1	109·6	17	43	·0005	23	·2782	91·1	·0059
1–2	3	98·0	104·0	26	33	·2175	33	·8184	89·1	·0068
2–3	4	98·0	103·9	22	36	·0435	28	·5168	54·3	·6135
3–4	5+	89·3	101·2	13	48	·0000	19	·2337	184·3	·0000
0–1	2–3	91·9	107·9	8	52	·0000	13	·2383	105·3	·0003
1–2	3–4	96·0	104·0	22	39	·0198	32	·8240	95·8	·0029
2–3	4+	87·2	101·3	10	51	·0000	17	·4674	201·4	·0000
0–1	2–4	89·3	107·0	6	54	·0000	11	·4268	131·3	·0000
1–2	3+	84·9	101·4	10	52	·0000	15	·1570	270·5	·0000
0–1	2+	77·1	102·0	1	60	·0000	3	1·0000	425·7	·0000
0–2	3	96·6	110·0	16	43	·0003	23	·3954	104·7	·0002
1–3	4	98·0	106·0	22	36	·0435	30	·7273	66·7	·2034
2–4	5+	88·8	101·8	12	49	·0000	15	·0335	264·6	·0000
0–2	3–4	93·8	109·0	10	51	·0000	18	·5745	132·6	·0000
1–3	4+	86·6	102·0	10	52	·0000	15	·1570	317·0	·0000
0–2	3+	80·9	102·7	3	59	·0000	7	1·0000	507·3	·0000
0–3	4	97·2	111·2	14	44	·0001	22	·5066	89·0	·0055
1–4	5+	88·1	102·5	10	52	·0000	15	·1570	370·2	·0000
0–3	4+	83·6	103·2	4	58	·0000	7	·1870	542·3	·0000
0–4	5+	85·8	103·7	5	57	·0000	11	1·0000	578·3	·0000

test or by the χ^2 test. Durations 1 and 2 differ by the signs test ($p(+)$ = ·0337), but not by the χ^2 test. The same is true for durations 2 and 4 ($p(+)$ = ·0314). Durations 2 and 3 reverse this, being different by the χ^2 test ($p(\chi^2)$ = ·0308), but not by the signs test. Durations 1 and 4 are very different according to the signs test ($p(+)$ = ·0003), but not by the χ^2 test. In no case does the runs test show a significant distinction.

When durations 1, 2, 3 and 4 are grouped similarly ambiguous results appear, with one test being passed and the other failed in almost every case. In none of the grouped comparisons shown are both tests passed, and in one (1–2 versus 3–4), both are failed. One may conclude that there is some evidence of a 'gradient' of increasing mortality with duration of selection over durations 1 to 4, but that it is not very conclusive. It would not be unreasonable to group all four durations together, nor would it be unreasonable to keep them separate.

Our preliminary investigation provides little justification for grouping 2–4 with 5+ for form 2+, as was done when the A1967-70 tables were constructed, but a comparison of this duration group with the previous graduation would be of interest. This is discussed in the accompanying CMI Report.

We therefore chose to graduate the following durations: 0, 1, 2, 3, 4, 2–4, 1–4, and 5+. This would allow select tables to be constructed:

0, 1, 2, 3, 4, 5+ (full five years selection)

0, 1, 2–4, 5+ (five years selection as A1967-70(5))

0, 1–4, 5+ (five years selection with only three groups)

We graduated each of these eight experiences by using 'central' exposures and graduating μ_x. A direct comparison with A1967-70 would be provided by using initial exposures and graduating q_x, as the CMI Committee had done at that time; this too is discussed in the accompanying Report.

17.5 *Durations 0 to 4*

We started with the relatively easy select durations, viz: 0, 1, 2, 3, 4, 2–4, and 1–4, seven sets of data in all.

The crude rates of μ_x for each age were calculated. Those for durations 0 and 2–4 are plotted in Figures 17.1 and 17.2. Confidence intervals for μ_x were also calculated, using the methods described in §2.6, and using the normal approximation for ages where the number of deaths exceeded 60 and the exact Poisson method for other ages. The upper and lower limits of the 95% confidence intervals for durations 0 and 2–4 are also shown in Figures 17.1 and 17.2.

The general shapes of the crude rates are very similar for all seven experiences. The sizes of the gates are also similar for the single durations (0 to 4), and they become rather narrower for duration groups 2–4 and 1–4. It is clear in every case that the crude rates start relatively high at the youngest ages for which any reasonable numbers are shown, falling to the late 20s of age, and rising almost linearly (on the vertical log scale) as age advances. These shapes are similar to that of the 1967-70 experiences, and suggest that a similar formula might suit.

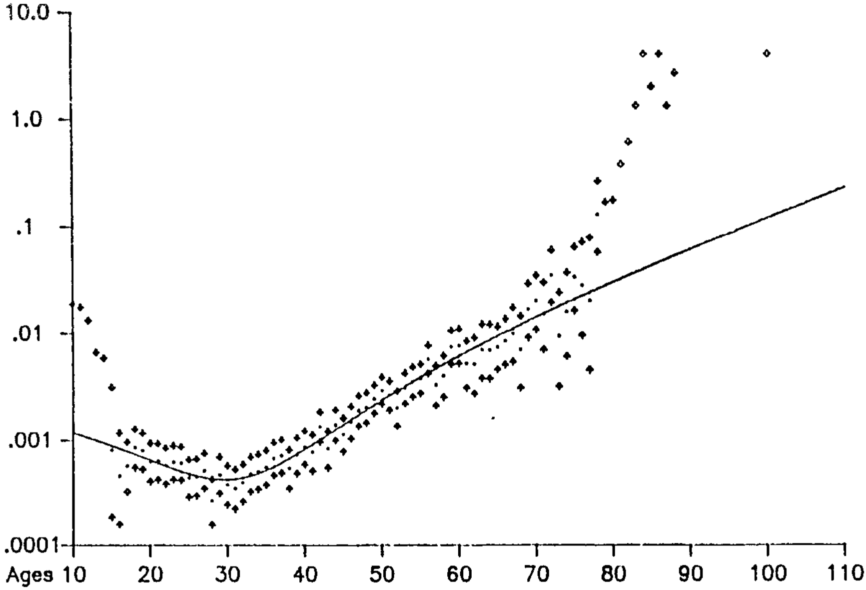

Figure 17.1. Male assured lives Duration 0 Crude rates and gates: $\mu_x = GM(2, 2)$.

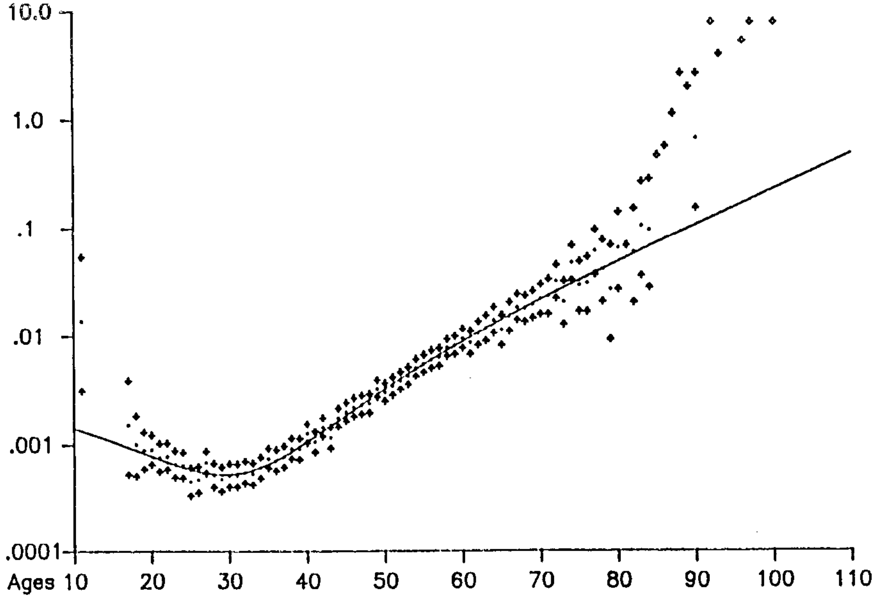

Figure 17.2. Male assured lives Durations 2–4 Crude rates and gates: $\mu_x = GM(2, 2)$.

On that occasion the CMI Committee used Barnett's formula, i.e. LGM(2, 2), graduating q_x with initial exposures.

17.6 *Choice of order of formula*

We explored as before which order of formula might be suitable, and as before began with graduations of μ_x, using a GM(r, s) formula, and a maximum likelihood (L_1) criterion, and with values of r and s with $s \geqslant 2$, $r + s \leqslant 6$, a total of 15 formulae. The values of the L_1 criterion (minus a large constant) for each duration are shown below.

Duration 0

r \ *s*	2	3	4	5	6
0	−126·8	−67·6	−27·7	−25·7	−25·2
1	−53·2	−36·4	−17·5	−17·3	
2	−18·0	−17·9	−16·9		
3	−18·0	−17·9			
4	−18·0				

Duration 1

r \ *s*	2	3	4	5	6
0	−145·6	−89·5	−25·3	−18·8	−12·7
1	−67·7	−35·2	−17·4	−12·0	
2	−12·3	−12·1	−10·7		
3	−12·0	−12·0			
4	−12·0				

Duration 2

r \ *s*	2	3	4	5	6
0	−82·9	−52·1	−13·1	−7·1	−6·9
1	−34·0	−17·1	−4·9	−3·4	
2	−3·8	−3·5	−3·5		
3	−3·5	−3·5			
4	−3·5				

Duration 3

r \ *s*	2	3	4	5	6
0	−87·3	−55·4	−8·8	−5·2	−5·1
1	−33·4	−13·3	−4·9	−4·9	
2	−5·5	−5·4	−4·6		
3	−5·5	−4·7			
4	−4·9				

Duration 4

	s 2	3	4	5	6
r					
0	−145·7	−129·3	−92·7	−85·1	−84·9
1	−112·3	−95·1	−77·7	−77·2	
2	−78·7	−77·5	−77·5		
3	−77·3	−76·9			
4	−77·3				

Durations 2–4

2	*s* 3	4	5	6	
r					
0	−318·9	−240·9	−119·0	−103·4	−103·4
1	−183·4	−129·7	−95·9	−92·8	
2	−94·0	−93·5	−93·2		
3	−93·3	−93·0			
4	−93·3				

Durations 1–4

2	*s* 3	4	5	6	
r					
0	−379·4	−244·3	−53·2	−29·7	−29·1
1	−160·0	−74·4	−52·1	−14·0	
2	−14·4	−14·1	−14·1		
3	−14·1	−14·1			
4	−14·1				

In every case a GM(2, 2) formula is the lowest order that one might stop at; any higher order formula hardly reduces the L_1 criterion at all, and certainly not significantly. We also observed that in most cases when we reached a formula with six parameters (i.e. GM(0, 6) to GM(4, 2)), the added parameters had very low *T*-ratios, i.e. they were not significantly different from zero. Going to a higher order formula would therefore be no better.

In every case the optimum GM(2, 2) formula proved to be a reasonably satisfactory fit. Tables 17.2 and 17.3 summarise the results for this formula for each duration group. Tables 17.4 and 17.5 give the detailed results for each age for duration 0 and durations 2–4. The results for the other durations are similar. Figures 17.1 and 17.2 show the values of μ_x according to the GM(2, 2) formula for Durations 0 and 2–4, and Figures 17.3 and 17.4 show the corresponding sheaves.

17.7 *Comparison of graduation formulae*

In § 13.5 we discussed how, given two sets of data, to each of which has been fitted the same order of formula, the maximum likelihood estimates of the two sets of parameters can be compared by calculating the squared 'distance' between the points in the parameter space. The measure

$$D = (\hat{\alpha}^{\mathrm{I}} - \hat{\alpha}^{\mathrm{II}})'(V^{\mathrm{I}} + V^{\mathrm{II}})^{-1}(\hat{\alpha}^{\mathrm{I}} - \hat{\alpha}^{\mathrm{II}})$$

is distributed as $\chi^2(n)$, where n is the number of parameters.

We have five independent sets of data and two which are formed by groups of the independent sets. We have fitted the same GM(2, 2) to each of them, and we have available the parameter estimates and the variance-covariance matrix for each of them. We are therefore able to calculate the measure, D, for each pair of durations or duration groups. The values of D and the values of $p(D)$, the probability that a value of D as great or greater would be obtained from a $\chi^2(4)$ distribution, are given below.

Value of D

Duration	1	2	3	4	2–4	1–4
Duration						
0	24·82	59·23	89·88	59·06	125·56	75·47
1		6·46	10·55	19·56	14·92	9·30
2			3·05	5·07	3·50	1·86
3				3·96	1·60	3·39
4					4·15	6·28
2–4						1·97

Value of $p(D)$

Duration	1	2	3	4	2–4	1–4
Duration						
0	·0001	·0000	·0000	·0000	·0000	·0000
1		·1675	·0321	·0006	·0049	·0540
2			·5494	·2801	·4782	·7608
3				·4115	·8083	·4941
4					·3856	·1795
2–4						·7418

Duration 0 is clearly different from the others, as we had already observed when comparing the crude data. Duration 1 is not significantly different from duration 2, but can be distinguished from durations 3 and 4. Durations 2, 3 and 4 are not significantly different from one another, so grouping them to form durations 2–4 is justified. Duration 1 is different from durations 2–4, so it is justifiable to leave it separate. Comparisons such as, for example, duration 2 with durations 2–4 and 1–4 are strictly invalid, since one duration is included in the other, and the data sets being compared are therefore not independent.

This test of the parameters of the graduated rates appears to be a more powerful discriminator between different levels of mortality experience than the preliminary tests of the crude data discussed in § 17.4 above. We conclude that durations 0, 1 and 2–4 should be left separate, the same as in the A1967-70(5) tables.

Inspection of the values of q_x for these three durations (Tables 17.2 and 17.3) shows that they are reasonably 'parallel' from age 30 onwards, but that the rates for duration 1 are above those for durations 2–4 at age 20.

The values of q_x actually cross over between ages 24 and 25. Whether this is a feature that should be eliminated in practical tables is a matter for consideration.

17.8 *Durations 5 and over*

We now turn to the ultimate experience, duration 5+. We first divided the exposed to risk and numbers of deaths by the variance ratios for each age derived from the count of duplicates among the deaths from this experience. The original numbers, the variance ratios, and the adjusted numbers are shown in Table 17.6, and the adjusted numbers are used in Tables 17.9 and 17.10.

The values of the crude rates, μ_x, and of the 95% confidence intervals, the gates, based on the adjusted numbers, are plotted in Figure 17.5. (Note that the adjustment by variance ratios does not alter the values of the crude rates, A_x/R_x.) It can be seen that the general pattern of the crude rates and of the gates for most of the age range is similar to that of the select durations, and that the gates are generally much narrower. But at the most advanced ages, above about 90, a different pattern emerges. The crude rates for ages 92 to 96 drop below the earlier linear trend, and from ages 97 to 103 are generally below the level reached by age 96. The gates, although widening, are sufficiently narrow to justify either a falling or at least a level trend of mortality with advancing age, if the data can be relied on.

The CMI Committee observed the same feature when investigating the corresponding 1967-70 experience. It seems very unlikely that lives assured at advanced ages should be so rejuvenated that their mortality rates contradict every other known experience, and we must look for an explanation in the possibility that some policies remain in the in force records of life offices, even though the life assured is dead. The instructions of the CMI Bureau to offices ask them to remove from the investigation at least multiple policies where there are no further premiums due and no bonus notices sent out, so that contact with the policyholder may be lost. But even if offices are able to do this reliably, policies so excluded may have contributed to the exposed to risk before being removed from the investigation. Thus the exposed to risk may be inflated invalidly.

To fit a satisfactory curve to this erratic part of the data would require a rather high order formula, in which mortality rates dropped after about age 96. We do not think that this would produce a practical mortality table. Further, the presence of this dubious data could distort any curve fitted to the main part of the table. We therefore followed the previous example, and omitted ages above 90 from the data when fitting a formula.

We then tried a number of formulae, fitting μ_x with a GM(r, s) formula, up to $r + s = 11$. The resulting values of the L_1 criterion (plus a large constant) are shown below.

r \ s	2	3	4	5	6	7	8	9	10	11
0	−293·4	−278·1	−156·1	−73·6	−47·0	−24·7	−20·1	−17·9	−17·7	−15·9
1	−277·7	−94·5	−64·8	−31·7	−30·8	−24·1	−18·6	−17·9	−17·6	
2	−37·5	−37·4	−35·6	−30·0	−24·7	−23·4	−17·3	−17·0		
3	−37·5	−36·7	−35·6	−19·3	−17·0	−17·0	−17·0			
4	−37·2	−33·8	−27·6	−18·0	−17·0	−17·0				
5	−37·2	−20·4	−17·5	−17·5	−17·0					
6	−37·2	−20·4	−17·4	−17·4						
7	−36·9	−19·0	−16·6							
8	−32·8	−16·9								
9	−32·2									

As we moved on to higher orders of formula we discovered that multiple maxima appear to exist. It is difficult to search an n-parameter space to establish this, when n exceeds about 4. The maximum value of the criterion (L_1 in this case) for any GM(r, s) formula must not be less than the values of the criterion both for the GM($r - 1$, s) and the GM(r, $s - 1$) formulae, if they exist. We therefore had to arrange that the search for the maximum point for each GM(r, s) formula commenced at the point represented by the parameters of whichever of GM($r - 1$, s) and GM(r, $s - 1$) had the higher value of the criterion.

Table 17.2. *Male assured lives, permanent, U.K., Durations 0, 1, 2, 3, Statistics for graduations with GM(2, 2) formula*
Function: μ_x, criterion: maximum likelihood

Duration	0	1	2	3
Values of parameters at optimum point:				
$a_0 \times 100$	−·465192	−·713368	−·473123	−·394346
(standard error)	·131830	·177278	·121012	·099159
T-ratio	−3·53	−4·02	−3·91	−3·98
$a_1 \times 100$	−·452546	−·676049	−·474513	−·423833
(standard error)	·102930	·135932	·101693	·089128
T-ratio	−4·40	−4·97	−4·67	−4·76
b_0	−3·985723	−3·689744	−3·687542	−3·620979
(standard error)	·051202	·050065	·039721	·038291
T-ratio	−77·84	−73·70	−92·84	−94·56
b_1	3·185063	3·027036	3·573205	3·989363
(standard error)	·387967	·340837	·328751	·313282
T-ratio	8·21	8·88	10·87	12·73
(Ages grouped so that each $E_x \geqslant 5$):				
Signs test:				
Number of +	30	28	31	30
Number of −	30	32	29	30
p(pos)	·5513	·3494	·6506	·5513
Runs test:				
No. of runs	30	28	37	39
p(runs)	·4491	·2686	·9560	·9869

Table 17.2 (*cont.*)

Duration	0	1	2	3
Kolmogorov–Smirnov test:				
Max deviation	·0105	·0097	·0079	·0090
p(KS)	1·0000	·9999	1·0000	1·0000
Serial Correlation test:				
r_1	−·0866	·1662	−·1077	−·2440
T-ratio	−·67	1·29	−·83	−1·89
r_2	−·0591	·0211	−·3756	·0981
T-ratio	−·46	·16	−2·91	·76
r_3	·0658	−·1378	·0711	−1886
T-ratio	·51	−1·07	·55	−1·46
χ^2 test				
χ^2	71·14	57·51	66·20	66·00
Deg. of freedom	56	56	56	56
$p(\chi^2)$	·0837	·4190	·1652	·1696

Specimen values of q_x for integral ages, simulated standard errors, and standard errors as percentage of value of q_x:

Age 20	·000622	·000807	·000695	·000768
(standard error)	·000142	·000162	·000099	·000083
percentage s.e.	22·83	20·05	14·28	10·83
Age 30	·000424	·000493	·000506	·000550
(standard error)	·000145	·000161	·000087	·000550
percentage s.e.	34·19	32·75	17·20	10·04
Age 40	·000855	·001042	·001109	·001099
(standard error)	·000161	·000178	·000106	·000073
percentage s.e.	18·79	17·0	9·56	6·62
Age 50	·002476	·003170	·003328	·003351
(standard error)	·000174	·000193	·000142	·000120
percentage s.e.	7·03	6·07	4·27	3·57
Age 60	·006333	·008173	·008830	·009359
(standard error)	·000287	·000315	·000312	·000311
percentage s.e.	4·54	3·86	3·53	3·32
Age 70	·014382	·018378	·020948	·023584
(standard error)	·001034	·001109	·001231	·001317
percentage s.e.	7·19	6·03	5·87	5·58
Age 80	·030205	·037884	·046198	·055427
(standard error)	·003390	·003535	·004324	·004966
Perc. s.e.	11·22	9·33	9·36	8·96
Age 90	·060190	·073663	·096683	·123387
(standard error)	·009467	·009558	·012744	·015501
percentage s.e.	15·73	12·98	13·18	12·56
Age 100	·115091	·136765	·192489	·258103
(standard error)	·023357	·022701	·032102	·039914
percentage s.e.	20·29	16·60	16·68	15·46
Age 110	·210921	·242281	·358468	·488360
(standard error)	·051003	·047471	·066994	·078161
percentage s.e.	24·18	19·59	18·69	16·00

Table 17.3. *Male assured lives, permanent, U.K., Durations 4, 2–4, 1–4 Statistics for graduations with GM(2,2) formula Function: μ_x, criterion: maximum likelihood*

Duration	4	2–4	1–4
Values of parameters at optimum point:			
$a_0 \times 100$	−·701247	−·487122	−·539786
(standard error)	·186797	·071283	·066340
T-ratio	−3·80	−6·83	−8·14
$a_1 \times 100$	−·687857	−·496613	−·540869
(standard error)	·152356	·061182	·055425
T-ratio	−4·51	−8·12	−9·76
b_0	−3·564739	−3·634119	−3·648090
(standard error)	·048345	·022500	·019942
T-ratio	−73·74	−161·52	−182·94
b_1	3·227393	3·647534	3·494553
(standard error)	·337559	·186707	·162494
T-ratio	9·56	19·54	21·51
(Ages grouped so that each $E_x \geqslant 5$):			
Signs test:			
Number of +	30	29	31
Number of −	29	36	35
p(pos)	·6026	·2285	·3561
Runs test:			
Number of runs	34	34	38
p(runs)	·8544	·6359	·8755
Kolmogorov–Smirnov test:			
Max deviation	·0101	·0045	·0034
p(KS)	·9993	1·0000	1·0000
Serial Correlation test:			
$r1$	−·0739	−·3213	−·1532
T-ratio	−·57	−2·59	−1·24
$r2$	−·0524	−·0123	−·0739
T-ratio	−·40	−·10	−·60
$r3$	−·1690	−·0985	−·1255
T-ratio	−1·30	−·79	−1·02
χ^2 test			
χ^2	58·94	73·30	85·12
Degrees of freedom	55	61	62
$p(\chi^2)$	·3336	·1344	·0274
Specimen values of q_x for integral ages, simulated standard errors, and standard errors as percentage of value of q_x:			
Age 20	·000867	·000759	·000775
(standard error)	·000186	·000046	·000038
percentage s.e.	21·45	6·03	4·86
Age 30	·000543	·000532	·000522
(standard error)	·000173	·000028	·000023
percentage s.e.	31·96	5·18	4·46

Table 17.3 (*cont.*)

Duration	4	2–4	1–4
Age 40	·001172	·001129	·001106
(standard error)	·000195	·000037	·000031
percentage s.e.	16·60	3·27	2·82
Age 50	·003614	·003428	·003372
(standard error)	·000216	·000066	·000056
percentage s.e.	5·98	1·92	1·66
Age 60	·009492	·009237	·008998
(standard error)	·000324	·000177	·000153
percentage s.e.	3·42	1·91	1·70
Age 70	·021825	·022223	·021291
(standard error)	·001196	·000723	·000609
percentage s.e.	5·48	3·25	2·86
Age 80	·046110	·049629	·046602
(standard error)	·004120	·002605	·002127
percentage s.e.	8·93	5·25	4·57
Age 90	·091894	·105003	·096541
(standard error)	·011730	·007809	·006223
percentage s.e.	12·76	7·44	6·45
Age 100	·174228	·210623	·190101
(standard error)	·028641	·019782	·015558
percentage s.e.	16·44	9·39	8·18
Age 110	·311969	·392273	·350689
(standard error)	·059491	·040890	·032524
percentage s.e.	19·07	10·42	9·27

For example, in the table above, when we commenced the search for the parameters for the GM(2, 4) formula, we looked at its neighbours above and to the left, GM(1, 4) and GM(2, 3) respectively. The values of L_1 for these two formulae are $-64{\cdot}8$ and $-37{\cdot}4$ respectively. The higher of these is $-37{\cdot}4$, given by the GM(2, 3) formula. We therefore began the search for GM(2, 4) at the point given by the optimum parameters for the GM(2, 3) formula, with the additional parameter b_4 taken to be zero. If we had begun at some other point the optimisation routines we have used might have found a local maximum that gave a value of L_1 less than $-37{\cdot}4$, which we know would not be the global maximum for GM(2, 4). We cannot be sure that we have in fact found global maxima everywhere, but our results are not inconsistent among themselves.

A summary of the characteristics of each of these graduations is shown in Table 17.7. Because the question of whether a particular formula provides a 'good fit' or not is more doubtful in this case, we give the values of p(runs) p(KS) and we indicate which of the first three serial correlation coefficients have T-ratios which are greater than 1·96. As it happens none is less than $-1{\cdot}96$.

Table 17.4. *Male assured lives, permanent, U.K. Duration 0. Details of graduation Aa1. Function: μ_x, GM(2, 2), criterion: maximum likelihood. Ages grouped so that each $E_x \geqslant 5$.*

Age	R_x	μ_x	A_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
10	208·7	·00118518	0	·25	−·25			
11	222·7	·00112142	0	·25	−·25			
12	295·2	·00105941	0	·31	−·31			
13	588·7	·00099927	0	·59	−·59			
14	665·2	·00094113	0	·63	−·63			
15	1242·5	·00088512	1	1·10	−·10			
16	6539·2	·00083139	3	5·44	−2·44			
10–16	9762·5		4	8·56	−4·56	2·93	−1·56	46·7
17	24585·5	·00078006	14	19·18	−5·18	4·38	−1·18	73·0
18	30739·0	·00073132	26	22·48	3·52	4·74	·74	115·7
19	37844·5	·00068533	30	25·94	4·06	5·09	·80	115·7
20	43326·7	·00064226	27	27·83	−·83	5·28	−·16	97·0
21	49138·7	·00060231	31	29·60	1·40	5·44	·26	104·7
22	55786·2	·00056569	32	31·56	·44	5·62	·08	101·4
23	60102·2	·00053261	37	32·01	4·99	5·66	·88	115·6
24	62813·2	·00050331	38	31·61	6·39	5·62	1·14	120·2
25	64179·2	·00047804	28	30·68	−2·68	5·54	−·48	91·3
26	65263·5	·00045705	29	29·83	−·83	5·46	−·15	97·2
27	64451·5	·00044064	33	28·40	4·60	5·33	·86	116·2
28	64758·5	·00042911	17	27·79	−10·79	5·27	−2·05	61·2
29	64444·2	·00042276	30	27·24	2·76	5·22	·53	110·1
30	66706·2	·00042196	25	28·15	−3·15	5·31	−·59	88·8
31	69428·7	·00042705	24	29·65	−5·65	5·45	−1·04	81·0
32	71226·5	·00043843	28	31·23	−3·23	5·59	−·58	89·7
33	71684·0	·00045652	34	32·73	1·27	5·72	·22	103·9
34	67905·2	·00048175	34	32·71	1·29	5·72	·23	103·9
35	63846·7	·00051458	35	32·85	2·15	5·73	·37	106·5
36	58683·0	·00055554	39	32·60	6·40	5·71	1·12	119·6
37	53938·0	·00060514	38	32·64	5·36	5·71	·94	116·4
38	48587·5	·00066395	26	32·26	−6·26	5·68	−1·10	80·6
39	44699·7	·00073259	32	32·75	−·75	5·72	−·13	97·7
40	42337·2	·00081169	36	34·36	1·64	5·86	·28	104·8
41	39369·7	·00090195	30	35·51	−5·51	5·96	−·92	84·5
42	36762·5	·00100410	49	36·91	12·09	6·08	1·99	132·7
43	35261·0	·00111892	29	39·45	−10·45	6·28	−1·66	73·5
44	33190·5	·00124725	46	41·40	4·60	6·43	·72	111·1
45	31962·2	·00138997	36	44·43	−8·43	6·67	−1·26	81·0
46	28744·2	·00154803	42	44·50	−2·50	6·67	−·37	94·4
47	26603·7	·00172244	50	45·82	4·18	6·77	·62	109·1
48	24809·5	·00191427	50	47·49	2·51	6·89	·36	105·3
49	23961·7	·00212468	58	50·91	7·09	7·14	·99	113·9
50	23864·5	·00235488	69	56·20	12·80	7·50	1·71	122·8
51	21059·5	·00260617	55	54·88	·12	7·41	·02	100·2

Table 17.4 (*cont.*)

Age	R_x	μ_x	A_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
52	17616·5	·00287994	35	50·73	−15·73	7·12	−2·21	69·0
53	16093·7	·00317768	49	51·14	−2·14	7·15	−·30	95·8
54	14846·5	·00350095	52	51·98	·02	7·21	·00	100·0
55	14686·5	·00385144	55	56·56	−1·56	7·52	−·21	97·2
56	11385·2	·00423093	64	48·17	15·83	6·94	2·28	132·9
57	7664·5	·00464134	25	35·57	−10·57	5·96	−1·77	70·3
58	5771·0	·00508469	23	29·34	−6·34	5·42	−1·17	78·4
59	5031·7	·00556316	37	27·99	9·01	5·29	1·70	132·2
60	4762·7	·00607905	36	28·95	7·05	5·38	1·31	124·3
61	3282·7	·00663483	17	21·78	−4·78	4·67	−1·02	78·0
62	2166·5	·00723312	11	15·67	−4·67	3·96	−1·18	70·2
63	1734·2	·00787671	12	13·66	−1·66	3·70	−·45	87·8
64	1747·0	·00856859	12	14·97	−2·97	3·87	−·77	80·2
65	2874·0	·00931192	21	26·76	−5·76	5·17	−1·11	78·5
66	2157·2	·01011010	18	21·81	−3·81	4·67	−·82	82·5
67	1225·5	·01096674	12	13·44	−1·44	3·67	−·39	89·3
68	865·7	·01188567	6	10·29	−4·29	3·21	−1·34	58·3
69	729·2	·01287100	12	9·39	2·61	3·06	·85	127·8
70	613·2	·01392709	12	8·54	3·46	2·92	1·18	140·5
71	471·0	·01505860	7	7·09	−·09	2·66	−·03	98·7
72	380·5	·01627048	13	6·19	6·81	2·49	2·74	210·0
73	324·5	·01756802	3	5·70	−2·70	2·39	−1·13	52·6
74	256·0	·01895687	4	4·85	−·85			
75	243·0	·02044301	8	4·97	3·03			
74–75	499·0		12	9·82	2·18	3·13	·70	122·2
76	109·3	·02203286	3	2·41	·59			
77	50·8	·02373324	1	1·20	−·20			
78	47·5	·02555141	6	1·21	4·79			
79	23·5	·02749511	0	·65	−·65			
80	22·5	·02957262	0	·67	−·67			
81	10·5	·03179273	0	·33	−·33			
82	6·5	·03416481	0	·22	−·22			
83	3·0	·03669887	0	·11	−·11			
84	1·0	·03940556	0	·04	−·04			
85	2·0	·04229623	0	·08	−·08			
86	1·0	·04538298	0	·05	−·05			
87	3·0	·04867871	0	·15	−·15			
88	1·5	·05219717	0	·08	−·08			
89–99	no data							
100	1·0	·11822877	0	·12	−·12			
76–100	283·0		10	7·31	2·69	2·70	·99	136·7
T. 1799039·7			1795	1795·02	−·02			100·0

$\chi^2 = \Sigma z_x^2 = 71{\cdot}14$

Table 17.5. *Male assured lives, permanent, U.K., Durations 2–4. Details of graduation Aa1. Function: μ_x, GM(2, 2), criterion: maximum likelihood. Ages grouped so that each $E_x \geqslant 5$*

Age	R_x	μ_x	A_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
10	410·7	·00141988	0	·58	−58			
11	72·3	·00134566	1	·10	·90			
12	160·2	·00127335	0	·20	−·20			
13	474·7	·00120307	0	·57	−·57			
14	834·2	·00113500	0	·95	−·95			
15	1160·0	·00106929	0	1·24	−1·24			
16	1328·5	·00100612	0	1·34	−1·34			
17	1950·2	·00094569	3	1·84	1·16			
10–17	6390·9		4	6·82	−2·82	2·61	−1·08	58·6
18	8899·5	·00088820	9	7·90	1·10	2·81	·39	113·9
19	33660·0	·00083388	30	28·07	1·93	5·30	·36	106·9
20	59836·5	·00078296	54	46·85	7·15	6·84	1·04	115·3
21	84059·0	·00073571	64	61·84	2·16	7·86	·27	103·5
22	96024·0	·00069240	75	66·49	8·51	8·15	1·04	112·8
23	110277·7	·00065332	73	72·05	·95	8·49	·11	101·3
24	125642·5	·00061881	81	77·75	3·25	8·82	·37	104·2
25	141821·7	·00058920	64	83·56	−19·56	9·14	−2·14	76·6
26	158482·2	·00056486	75	89·52	−14·52	9·46	−1·53	83·8
27	171814·5	·00054621	118	93·85	24·15	9·69	2·49	125·7
28	182738·2	·00053365	95	97·52	−2·52	9·88	−·26	97·4
29	190361·0	·00052766	91	100·45	−9·45	10·02	−·94	90·6
30	196849·7	·00052874	102	104·08	−2·08	10·20	−·20	98·0
31	204916·2	·00053741	106	110·12	−4·12	10·49	−·39	96·3
32	218434·7	·00055426	121	121·07	−·07	11·00	−·01	99·9
33	224663·0	·00057989	120	130·28	−10·28	11·41	−·90	92·1
34	218113·5	·00061499	133	134·14	−1·14	11·58	−·10	99·2
35	206441·7	·00066025	154	136·30	17·70	11·67	1·52	113·0
36	188728·7	·00071646	135	135·22	−·22	11·63	−·02	99·8
37	172812·2	·00078444	134	135·56	−1·56	11·64	−·13	98·8
38	159564·2	·00086508	147	138·04	8·96	11·75	·76	106·5
39	145668·0	·00095934	133	139·74	−6·74	11·82	−·57	95·2
40	134140·0	·00106824	171	143·29	27·71	11·97	2·31	119·3
41	126663·2	·00119291	135	151·10	−16·10	12·29	−1·31	89·4
42	122415·0	·00133453	178	163·37	14·63	12·78	1·14	109·0
43	117218·2	·00149438	136	175·17	−39·17	13·24	−2·96	77·6
44	109732·7	·00167384	196	183·68	12·32	13·55	·91	106·7
45	100713·5	·00187441	203	188·78	14·22	13·74	1·04	107·5
46	92700·2	·00209766	206	194·45	11·55	13·94	·83	105·9
47	87354·2	·00234533	204	204·87	−·87	14·31	−·06	99·6
48	82443·5	·00261926	198	215·94	−17·94	14·69	−1·22	91·7
49	77393·0	·00292143	253	226·10	26·90	15·04	1·79	111·9
50	71289·0	·00325399	217	231·97	−14·97	15·23	−·98	93·5
51	65584·0	·00361924	226	237·36	−11·36	15·41	−·74	95·2
52	62054·5	·00401964	240	249·44	−9·44	15·79	−·60	96·2
53	57247·2	·00445786	245	255·20	−10·20	15·97	−·64	96·0
54	51081·0	·00493675	258	252·17	5·83	15·88	·37	102·3
55	43579·2	·00545941	240	237·92	2·08	15·42	·14	100·9

Table 17.5 (*cont.*)

Age	R_x	μ_x	A_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
56	38207·0	·00602914	231	230·36	·64	15·18	·04	100·3
57	35992·0	·00664949	228	239·33	−11·33	15·47	−·73	95·3
58	32170·0	·00732431	248	235·62	12·38	15·35	·81	105·2
59	26721·2	·00805772	217	215·31	1·69	14·67	·12	100·8
60	18951·2	·00885415	177	167·80	9·20	12·95	·71	105·5
61	12914·5	·00971836	111	125·51	−14·51	11·20	−1·29	88·4
62	9769·7	·01065549	102	104·10	−2·10	10·20	−·21	98·0
63	7476·5	·01167106	87	87·26	−·26	9·34	−·03	99·7
64	5801·2	·01277100	80	74·09	5·91	8·61	·69	108·0
65	4530·2	·01396170	51	63·25	−12·25	7·95	−1·54	80·6
66	3821·7	·01525002	57	58·28	−1·28	7·63	−·17	97·8
67	4393·2	·01664336	80	73·12	6·88	8·55	·80	109·4
68	4473·0	·01814965	79	81·18	−2·18	9·01	−·24	97·3
69	3790·5	·01977746	73	74·97	−1·97	8·66	−·23	97·4
70	2441·5	·02153597	53	52·58	·42	7·25	·06	100·8
71	1557·0	·02343507	36	36·49	−·49	6·04	−·08	98·7
72	1256·7	·02548542	40	32·03	7·97	5·66	1·41	124·9
73	1026·7	·02769844	21	28·44	−7·44	5·33	−1·40	73·8
74	760·5	·03008646	36	22·88	13·12	4·78	2·74	157·3
75	513·7	·03266271	15	16·78	−1·78	4·10	−·43	89·4
76	391·7	·03544144	12	13·88	−1·88	3·73	−·51	86·4
77	314·5	·03843798	19	12·09	6·91	3·48	1·99	157·2
78	219·2	·04166880	9	9·14	−·14	3·02	−·04	98·5
79	111·8	·04515165	3	5·05	−2·05	2·25	−·91	59·5
80	77·5	·04890558	5	3·79	1·21			
81	57·3	·05295111	0	3·03	−3·03			
80–81	134·7		5	6·82	−1·82	2·61	−·70	73·3
82	50·8	·05731032	3	2·91	·09			
83	28·8	·06200694	3	1·78	1·22			
84	21·0	·06706650	2	1·41	·59			
85	8·5	·07251647	0	·62	−·62			
86	7·0	·07838640	0	·55	−·55			
87	3·5	·08470807	0	·30	−·30			
88	1·5	·09151566	0	·14	−·14			
89	2·0	·09884595	0	·20	−·20			
90	1·5	·10673850	1	·16	·84			
91	no data							
92	0·5	·12438377	0	·06	−·06			
93	1·0	·13423151	0	·13	−·13			
94	no data							
95	·3	·15624224	0	·04	−·04			
96	·8	·16852349	0	·13	−·13			
97	·5	·18174168	0	·09	−·09			
98–99	no data							
100	0·5	·22775409	0	·11	−·11			
82–100	128·0		9	8·62	·38	2·94	·13	104·4
T.	4925676·8		7503	7503·00	·00			100·00

$\chi^2 = \Sigma z_x^2 = 73{\cdot}30$

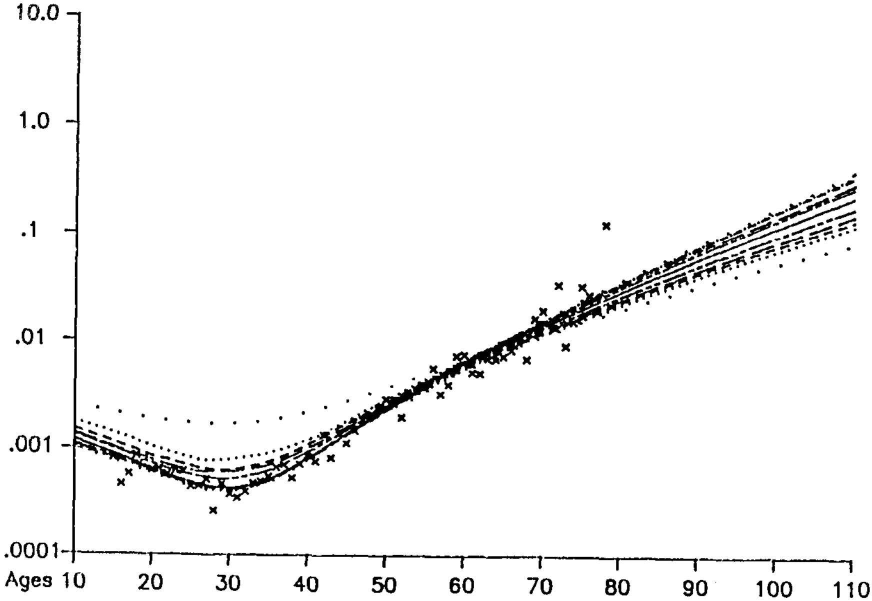

Figure 17.3. Male assured lives Duration 0. Sheaf for μ_x = GM(2, 2).

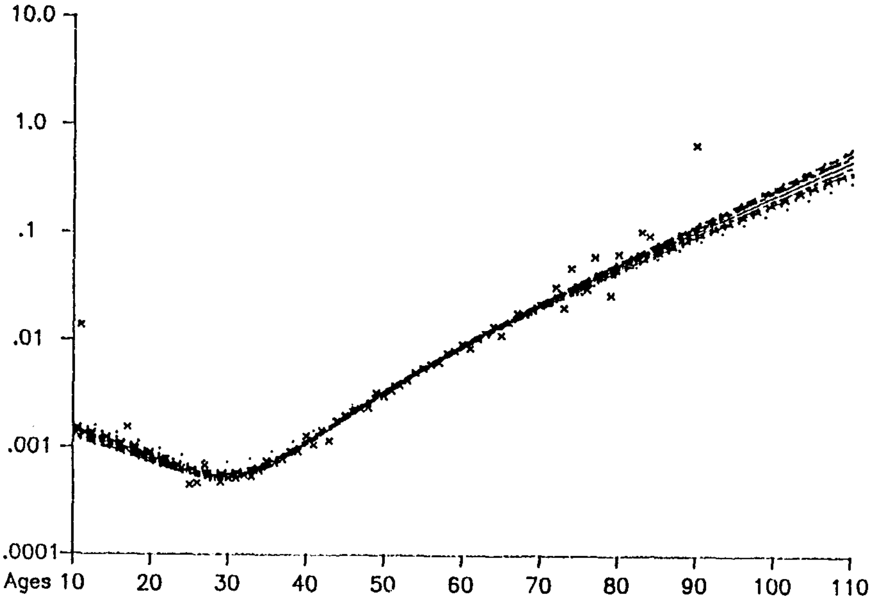

Figure 17.4. Male assured lives Durations 2–4. Sheaf for μ_x = GM(2,2).

Table 17.6. *Male assured lives (U.K. Permanent). Duration 5+*
Original data, variance ratios and adjusted data

Age	Original data R_x	Original data A_x	Variance ratio r_x	Adjusted data R'_x	Adjusted data A'_x
10	117·0	0	1·00	117·00	·00
11	33·5	0	1·00	33·49	·00
12	49·7	0	1·00	49·74	·00
13	56·8	0	1·00	56·75	·00
14	79·7	1	1·00	79·74	1·00
15	238·0	1	1·00	237·99	1·00
16	538·7	1	1·00	538·74	1·00
17	860·5	0	1·00	860·49	·00
18	1137·5	2	1·00	1137·49	2·00
19	1433·5	0	1·00	1433·50	·00
20	2330·0	3	1·00	2329·99	3·00
21	9149·5	4	1·67	5478·74	2·40
22	29954·0	29	1·21	24755·37	23·97
23	50989·7	46	1·12	45526·55	41·07
24	74643·0	52	1·00	74642·99	52·00
25	100433·0	73	1·18	85112·70	61·86
26	129037·0	81	1·34	96296·27	60·45
27	163196·8	95	1·22	133767·83	77·87
28	202341·0	99	1·27	159323·62	77·95
29	247811·5	153	1·31	189169·08	116·79
30	300754·5	160	1·26	238694·05	126·98
31	363992·8	186	1·23	295929·07	151·22
32	437952·7	247	1·54	284384·90	160·39
33	500565·5	338	1·60	312853·44	211·25
34	539011·8	343	1·35	399267·96	254·07
35	563449·0	431	1·55	363515·48	278·06
36	565514·0	398	1·54	367216·88	258·44
37	557513·0	446	1·48	376697·97	301·35
38	550015·2	537	1·82	302206·18	295·05
39	533139·2	480	1·68	317344·79	285·71
40	518725·0	533	1·58	328306·96	337·34
41	516972·2	649	1·72	300565·26	377·33
42	526998·2	719	1·80	292776·81	399·44
43	536894·2	771	1·79	299940·92	430·73
44	537019·8	939	1·71	314046·64	549·12
45	528910·8	1015	1·67	316713·02	607·78
46	520267·2	1146	1·87	278217·78	612·83
47	513590·7	1189	1·69	303899·85	703·55
48	509219·0	1327	1·74	292654·60	762·64
49	505844·5	1702	1·80	281024·72	945·56
50	498936·0	1868	1·82	274140·65	1026·37
51	489236·5	2095	1·74	281170·40	1204·02
52	477414·2	2242	1·80	265230·13	1245·56
53	465926·2	2508	1·66	280678·46	1510·84
54	453919·0	2656	1·65	275102·42	1609·70

Table 17.6 (*cont.*)

Age	Original data R_x	A_x	Variance ratio r_x	Adjusted data R'_x	A'_x
55	434168·7	2861	1·67	259981·28	1713·17
56	416652·2	3105	1·79	232766·62	1734·64
57	402193·2	3277	1·74	231145·55	1883·33
58	391892·8	3560	1·69	231889·20	2106·51
59	381495·5	4037	1·73	220517·63	2333·53
60	319429·5	3550	1·55	206083·55	2290·32
61	265730·0	3341	1·55	171438·71	2155·48
62	230875·0	3311	1·55	148951·61	2136·13
63	201601·0	3265	1·46	138082·88	2236·30
64	179806·5	3016	1·53	117520·58	1971·24
65	100538·2	1876	1·37	73385·58	1369·34
66	56673·0	1115	1·25	45338·40	892·00
67	46895·0	1014	1·39	33737·40	729·50
68	40987·2	1033	1·48	27694·09	697·97
69	36769·8	1011	1·32	27855·87	765·91
70	32995·0	1022	1·32	24996·21	774·24
71	30145·5	1074	1·50	20097·00	716·00
72	28037·5	1054	1·47	19073·12	717·01
73	25884·5	1127	1·44	17975·34	782·64
74	23618·5	1125	1·41	16750·70	797·87
75	20875·0	1154	1·44	14496·53	801·39
76	18290·7	1098	1·44	12701·90	762·50
77	16143·5	1099	1·48	10907·77	742·57
78	14128·2	957	1·51	9356·46	633·77
79	12380·0	1025	1·46	8479·45	702·05
80	10255·0	967	1·43	7171·33	676·22
81	8607·0	875	1·59	5413·20	550·31
82	7304·8	762	1·40	5217·68	544·29
83	6173·2	752	1·33	4641·54	565·41
84	5035·2	648	1·39	3622·48	466·19
85	4203·2	545	1·41	2981·03	386·52
86	3480·7	508	1·35	2578·33	376·30
87	2925·2	453	1·27	2303·34	356·69
88	2384·5	394	1·22	1954·50	322·95
89	1955·0	384	1·32	1481·06	290·91
90	1622·0	314	1·36	1192·65	230·88
91	1307·8	284	1·33	983·27	213·53
92	1028·5	229	1·30	791·15	176·15
93	771·2	185	1·50	514·17	123·33
94	568·5	141	1·45	392·07	97·24
95	418·0	115	1·60	261·25	71·87
96	295·0	87	1·33	221·80	65·41
97	208·8	43	1·19	175·42	36·13
98	157·0	30	1·08	145·37	27·78
99	122·8	21	1·00	122·75	21·00
100	147·8	13	1·00	147·75	13·00
101	40·0	9	1·29	31·01	6·98
102	19·0	6	1·00	19·00	6·00

Table 17.6 (*cont.*)

Age	Original data R_x	A_x	Variance ratio r_x	Adjusted data R'_x	A'_x
103	10·0	1	1·00	10·00	1·00
104	4·5	0	1·00	4·50	·00
105	3·0	0	1·00	3·00	·00
106	1·0	0	1·00	1·00	·00
107	·0	0	1·00	·00	·00
108	4·5	0	1·00	4·50	·00
Tot.	17313471·2	83438		10827135·99	53239·27

An indication of which formulae have their final r-parameter and s-parameter both significant is shown below.

r \ s	2	3	4	5	6	7	8	9	10	11
0	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
1	Yes	Yes	Yes	Yes	No	No	Yes	No	No	
2	Yes	No	Yes	No	Yes	No	No	No		
3	No	No	No	Yes	Yes	No	No			
4	No	Yes	Yes	No	No	No				
5	No	Yes	No	No	No					
6	No	No	No	No						
7	No	No	No							
8	No	No								
9	No									

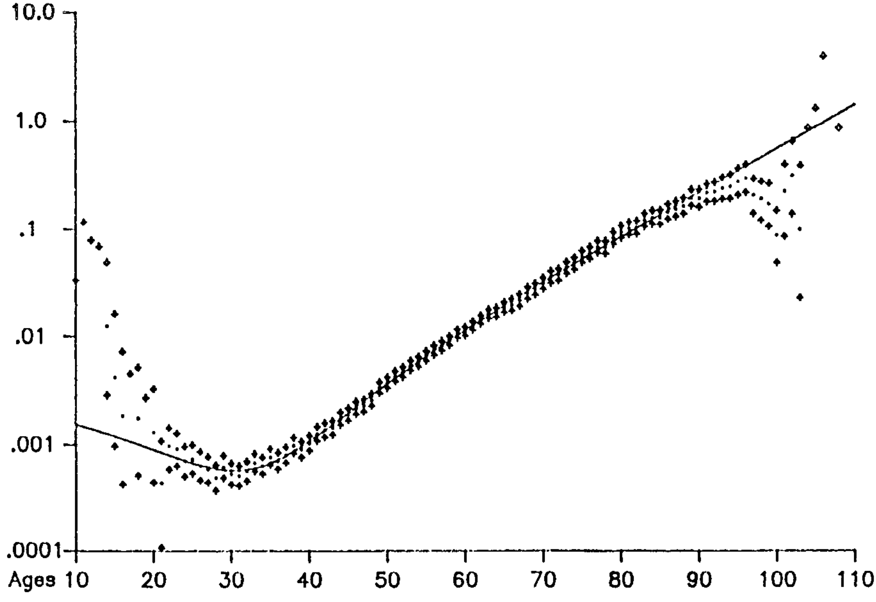

Figure 17.5. Male assured lives Duration 5+. Crude rates and gates: μ_x = GM(2, 2).

Table 17.7. *Male assured lives (U.K. Permanent) Duration 5+. Comparison of $GM(r, s)$ formulae for graduating μ_x*

Order (r, s)	Value of L_1	p (runs)	p (KS)	ser cor[(1)]	χ^2	$p(\chi^2)$	Sig. par.	Good shape[(2)]	Tight sheaf[(3)]
0, 2	−293·4	·00	·00	123	759·2	·000	Yes	No	Yes
0, 3	−278·1	·00	·00	123	800·5	·000	Yes	No	Yes
1, 2	−277·7	·00	·00	123	648·7	·000	Yes	No	?
0, 4	−156·1	·00	·00	123	377·5	·000	Yes	No	?
1, 3	−94·5	·00	·03	123	235·6	·000	Yes	No	Yes
2, 2	−37·5	·01	·85	1	109·0	·001	Yes	Yes	Yes
0, 5	−73·6	·00	·13	123	183·2	·000	Yes	?	?
1, 4	−·64·8	·04	·57	123	168·8	·000	Yes	No	Yes
2, 3	−37·4	·04	·80	1	108·9	·001	No	Yes	Yes
3, 2	−37·5	·01	·85	1	109·0	·001	No	Yes	?
0, 6	−47·0	·00	·47	123	124·2	·000	Yes	No	No
1, 5	−31·8	·45	·93	1	98·8	·003	Yes	No	No
2, 4	−35·6	·00	·57	1	106·1	·001	Yes	Yes	No
3, 3	−36·7	·05	·65	13	108·2	·001	No	Yes	No
4, 2	−37·2	·05	·74	13	109·3	·000	No	Yes	No
0, 7	−24·7	·32	1·00	–	85·0	·041	Yes	No	No
1, 6	−30·8	·64	·99	1	97·3	·004	No	No	No
2, 5	−30·0	·36	·82	–	94·3	·007	No	No	No
3, 4	−35·6	·00	·56	1	106·0	·001	No	Yes	No
4, 3	−33·8	·10	·71	1	101·2	·002	Yes	Yes	?
5, 2	−37·2	·05	·75	13	109·3	·000	No	Yes	No
0, 8	−20·1	·34	·99	–	75·2	·139	Yes	Yes	No
1, 7	−24·1	·33	·99	–	84·2	·038	No	No	No
2, 6	−24·7	·34	·96	–	83·7	·041	Yes	No	No
3, 5	−19·3	·52	1·00	–	73·9	·163	Yes	No	No
4, 4	−27·6	·24	·84	–	89·4	·016	Yes	No	?
5, 3	−20·4	·34	·98	–	76·2	·123	Yes	Yes	?
6, 2	−37·2	·05	·76	13	109·3	·000	No	Yes	No
0, 9	−17·9	·79	1·00	–	71·1	·200	Yes	No	No
1, 8	−18·6	·84	1·00	–	72·8	·163	Yes	?	No
2, 7	−23·4	·33	·99	–	82·4	·042	No	No	No
3, 6	−17·0	·52	1·00	–	69·4	·242	Yes	?	No
4, 5	−18·0	·33	1·00	–	71·6	·188	No	No	No
5, 4	−17·5	·89	1·00	–	70·2	·222	No	?	No
6, 3	−20·4	·33	·98	–	76·1	·107	No	Yes	No
7, 2	−36·9	·02	·84	1	109·0	·000	No	Yes	No
0, 10	−17·7	·78	1·00	–	70·9	·180	No	No	?
1, 9	−17·9	·63	1·00	–	71·1	·177	No	No	?
2, 8	−17·3	·70	1·00	–	69·9	·203	No	?	?
3, 7	−17·0	·52	1·00	–	69·4	·216	No	?	No
4, 6	−17·0	·52	1·00	–	69·4	·215	No	?	No
5, 5	−17·5	·89	1·00	–	70·2	·197	No	?	No
6, 4	−17·4	·89	1·00	–	70·6	·187	No	?	No
7, 3	−19·0	·65	1·00	–	73·5	·150	No	Yes	No
8, 2	−32·8	·10	·85	1	100·7	·001	No	Yes	No
0, 11	−15·9	·89	1·00	–	68·2	·218	No	No	No
1, 10	−17·6	·78	1·00	–	70·8	·161	No	No	No

Table 17.7 (*cont.*)

Order (r, s)	Value of L_1	p (runs)	p (KS)	ser cor[(1)]	χ^2	$p(\chi^2)$	Sig. par.	Good shape[(2)]	Tight sheaf[(3)]
2, 9	−17·0	·32	1·00	–	68·9	·202	No	?	No
3, 8	−17·0	·52	1·00	–	69·3	·191	No	?	No
4, 7	−17·0	·52	1·00	–	69·4	·190	No	?	No
5, 6	−17·0	·52	1·00	–	69·4	·191	No	?	No
6, 5	−17·4	·89	1·00	–	70·6	·164	No	?	No
7, 4	−16·6	·96	1·00	–	68·5	·211	No	?	No
8, 3	−16·9	·71	1·00	–	69·2	·196	No	?	No
9, 2	−32·2	·12	1·00	–	99·7	·016	No	Yes	No

Notes (1) The digits indicate which serial correlation coefficients have T-ratios that exceed 1·96. (2) ? indicates that μ_x rises very sharply at very high ages. (3) ? indicates a sheaf of intermediate tightness.

The following table shows the values of $p(\chi^2)$ for each formula.

r \ s	2	3	4	5	6	7	8	9	10	11
0	·000	·000	·000	·000	·000	·041	·139	·200	·180	·218
1	·000	·000	·000	·003	·004	·038	·163	·177	·161	
2	·001	·001	·001	·007	·041	·042	·203	·202		
3	·001	·001	·001	·163	·242	·216	·191			
4	·000	·002	·016	·188	·215	·190				
5	·000	·123	·222	·197	·191					
6	·000	·107	·187	·164						
7	·000	·150	·211							
8	·001	·196								
9	·001									

In Table 17.7 we classify the shape of some of the low order formulae as poor, because they do not recognize the higher mortality at the youngest ages. We query the shape when the values of μ_x rise too sharply at advanced ages. We query the sheaf when it is tight over most of the age range, only bursting out at the very ends of the range.

Of the low order formulae, GM(2, 2) has the highest value of L_1, and there is little improvement in the value of L_1 in the immediately neighbouring orders of formula. However, the GM(2, 2) formula has a significantly low number of runs, and the T-ratio of its first serial correlation coefficient is 2·67, significantly large. The value of χ^2 is also large, and $p(\chi^2)$ is only ·0007. Inspection of the detailed results show significantly low values of z_x (less than −2·0) at ages 47, 48, 66 and 67, and significantly high values (more than 2·0) at ages 35, 38 and 80. The highest absolute value of z_x is 2·65 (at age 80). The two pairs of neighbouring large negative values help to contribute towards the high value of the first serial correlation coefficient.

All the parameters of the GM(2, 2) formula are very significantly non-zero, the shape of the curve is satisfactory, and the sheaf is very tight. It therefore has many good features, even though the fit is not ideal.

When the order of the formula is increased the value of L_1 increases sharply. Many of the formulae with high values of L_1 have poor shapes or non-significant

Table 17.8. *Male assured lives, permanent, U.K., Duration 5 +. Statistics for graduations with various formulae. Function: μ_x, criterion: maximum likelihood*

Formula	GM(2, 2)	GM(0, 8)	GM(3, 6)
Value of criterion at optimum point:			
L_1	−285637·5	−285620·1	−285617·0
Values of parameters at optimum point:			
$a_0 \times 100$	−·378772	–	1·497015
(standard error)	·022451	–	·136683
T-ratio	−16·87	–	10·95
$a_1 \times 100$	−·431902	–	2·021917
(standard error)	·024536	–	·194400
T-ratio	−17·60	–	10·40
$a_3 \times 100$	–	–	·628555
(standard error)	–	–	·065005
T-ratio	–	–	9·67
b_0	−3·329023	−1·500793	−11·045274
(standard error)	·008608	3·801014	1·378611
T-ratio	−386·74	−·39	−8·01
b_1	4·595701	8·232312	24·092333
(standard error)	·042362	7·104652	9·155591
T-ratio	108·49	1·16	2·63
b_2	–	3·754954	−9·639479
(standard error)	–	5·782655	1·878418
T-ratio	–	·65	−5·13
b_3	–	2·583607	9·559850
(standard error)	–	4·059745	4·862410
T-ratio	–	·64	1·97
b_4	–	2·576161	−2·381718
(standard error)	–	2·408545	·525176
T-ratio	–	1·07	−4·54
b_5	–	1·350736	2·207383
(standard error)	–	1·169434	1·120549
T-ratio	–	1·16	1·97
b_6	–	·780707	–
(standard error)	–	·425578	–
T-ratio	–	1·83	–
b_7	–	·293572	–
(standard error)	–	·105665	–
T-ratio	–	2·78	–
(Ages grouped so that each $E_x \geqslant 5$):			
Signs test:			
Number of +	36	32	33
Number of −	34	39	38
p(pos)	·6399	·2383	·3177
Runs test:			
Number of runs	26	34	36
p(runs)	·0109	·3449	·5171

Table 17.8 (*cont.*)

Formula	GM(2, 2)	GM(0, 8)	GM(3, 6)
Kolmogorov–Smirnov test:			
Max deviation	·0038	·0028	·0022
p(KS)	·8491	·9872	·9995
Serial Correlation test:			
r_1	·3191	·0467	−·0286
T-ratio	2·67	·39	−·24
r_2	·2312	−·0833	−·1567
T-ratio	1·93	−·70	−1·32
r_3	·2245	·0166	−·0353
T-ratio	1·88	·14	−·30
χ^2 test			
χ^2	109·01	75·21	6·40
Degrees of freedom	60	63	62
$p(\chi^2)$	·0007	·1393	·2421
Specimen values of q_x for integral ages, simulated standard errors, and standard errors as percentage of value of q_x:			
Age 20	·000867	·001011	·000988
(standard error)	·000041	·000120	·000073
percentage s.e.	4·74	11·89	7·34
Age 30	·000574	·000567	·000557
(standard error)	·000015	·000014	·000018
percentage s.e.	2·66	2·41	3·17
Age 40	·001141	·001128	·001144
(standard error)	·000014	·000016	·000021
percentage s.e.	1·23	1·38	1·86
Age 50	·003859	·003891	·003881
(standard error)	·000024	·000034	·000048
percentage s.e.	·63	·88	1·23
Age 60	·011929	·011888	·011874
(standard error)	·000063	·000077	·000094
percentage s.e.	·53	·65	·79
Age 70	·033134	·032568	·032524
(standard error)	·000196	·000306	·000337
percentage s.e.	·59	·94	1·04
Age 80	·085510	·089476	·089660
(standard error)	·000756	·001053	·001135
percentage s.e.	·88	1·18	1·27
Age 90	·205722	·175560	·180145
(standard error)	·002646	·010149	·008852
percentage s.e.	1·29	5·78	4·91
Age 100	·442872	·216909	·342699
(standard error)	·006748	·225091	·233789
percentage s.e.	1·52	103·77	68·22
Age 110	·771310	·984405	·999945
(standard error)	·009306	·466081	·406370
percentage s.e.	1·21	47·35	40·64

Table 17.9. *Male assured lives, permanent, U.K., Duration 5+. Details of graduation Aa1. Function: μ_x, GM(2, 2), criterion: maximum likelihood. Ages grouped so that each $E_x \geqslant 5$. Age range 10–90 only*

Age	R'_x	μ_x	A'_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
10	117·0	·00153937	·00	·18	−·18			
11	33·5	·00146688	·00	·05	−·05			
12	49·7	·00139572	·00	·07	−·07			
13	56·8	·00132603	·00	·08	−·08			
14	79·7	·00125795	1·00	·10	·90			
15	238·0	·00119163	1·00	·28	·72			
16	538·7	·00112724	1·00	·61	·39			
17	860·5	·00106497	·00	·92	−·92			
18	1137·5	·00100502	2·00	1·14	·86			
19	1433·5	·00094761	·00	1·36	−1·36			
20	2330·0	·00089299	3·00	2·08	·92			
10–20	6874·9		8·00	6·86	1·14	2·62	·43	116·6
21	5478·7	·00084143	2·40	4·61	−2·21			
22	24755·4	·00079322	23·97	19·64	4·33			
21–22	30234·1		26·36	24·25	2·12	4·92	·43	108·7
23	45526·6	·00074869	41·07	34·09	6·99	5·84	1·20	120·5
24	74643·0	·00070818	52·00	52·86	−·86	7·27	−·12	98·4
25	85112·7	·00067210	61·86	57·20	4·66	7·56	·62	108·1
26	96296·3	·00064085	60·45	61·71	−1·26	7·86	−·16	98·0
27	133767·8	·00061491	77·87	82·26	−4·39	9·07	−·48	94·7
28	159323·6	·00059479	77·95	94·76	−16·81	9·73	−1·73	82·3
29	189169·1	·00058105	116·79	109·92	6·88	10·48	·66	106·3
30	238694·0	·00057430	126·98	137·08	−10·10	11·71	−·86	92·6
31	295929·1	·00057522	151·22	170·22	−19·00	13·05	−1·46	88·8
32	284384·9	·00058454	160·39	166·23	−5·85	12·89	−·45	96·5
33	312853·4	·00060308	211·25	188·68	22·57	13·74	1·64	112·0
34	399268·0	·00063171	254·07	252·22	1·85	15·88	·12	100·7
35	363515·5	·00067142	278·06	244·07	33·99	15·62	2·18	113·9
36	367216·9	·00072327	258·44	265·60	−7·16	16·30	−·44	97·3
37	376698·0	·00078843	301·35	297·00	4·35	17·23	·25	101·5
38	302206·2	·00086817	295·05	262·37	32·69	16·20	2·02	112·5
39	317344·8	·00096390	285·71	305·89	−20·18	17·49	−1·15	93·4
40	328307·0	·00107717	337·34	353·64	−16·30	18·81	−·87	95·4
41	300565·3	·00120966	377·33	363·58	13·74	19·07	·72	103·8
42	292776·8	·00136322	399·44	399·12	·32	19·98	·02	100·1
43	299940·9	·00153988	430·73	461·87	−31·15	21·49	−1·45	93·3
44	314046·6	·00174186	549·12	547·03	2·10	23·39	·09	100·4
45	316713·0	·00197160	607·78	624·43	−16·65	24·99	−67	97·3
46	278217·8	·00223178	612·83	620·92	−8·09	24·92	−·32	98·7
47	303899·8	·00252532	703·55	767·44	−63·89	27·70	−2·31	91·7
48	292654·6	·00285543	762·64	835·66	−73·01	28·91	−2·53	91·3
49	281024·7	·00322564	945·56	906·48	39·07	30·11	1·30	104·3
50	274140·7	·00363981	1026·37	997·82	28·55	31·59	·90	102·9
51	281170·4	·00410216	1204·02	1153·41	50·62	33·96	1·49	104·4
52	265230·1	·00461734	1245·56	1224·66	20·90	35·00	·60	101·7
53	280678·5	·00519044	1510·84	1456·84	54·00	38·17	1·41	103·7

Table 17.9 (*cont.*)

Age	R'_x	μ_x	A'_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
54	275102·4	·00582702	1609·70	1603·03	6·67	40·04	·17	100·4
55	259981·3	·00653320	1713·17	1698·51	14·66	41·21	·36	100·9
56	232766·6	·00731568	1734·64	1702·85	31·79	41·27	·77	101·9
57	231145·5	·00818181	1883·33	1891·19	−7·86	43·49	−·18	99·6
58	231889·2	·00913963	2106·51	2119·38	−12·87	46·04	−·28	99·4
59	220517·6	·01019799	2333·53	2248·84	84·69	47·42	1·79	103·8
60	206083·5	·01136654	2290·32	2342·46	−52·13	48·40	−1·08	97·8
61	171438·7	·01265591	2155·48	2169·71	−14·23	46·58	−·31	99·3
62	148951·6	·01407773	2136·13	2096·90	39·23	45·79	·86	101·9
63	138082·9	·01564474	2236·30	2160·27	76·03	46·48	1·64	103·5
64	117520·6	·01737092	1971·24	2041·44	−70·20	45·18	−1·55	96·6
65	73385·6	·01927160	1369·34	1414·26	−44·91	37·61	−1·19	96·8
66	45338·4	·02136357	892·00	968·59	−76·59	31·12	−2·46	92·1
67	33737·4	·02366525	729·50	798·40	−68·91	28·26	−2·44	91·4
68	27694·1	·02619684	697·97	725·50	−27·52	26·94	−1·02	96·2
69	27855·9	·02898046	765·91	807·28	−41·37	28·41	−1·46	94·9
70	24996·2	·03204037	774·24	800·89	−26·65	28·30	−·94	96·7
71	20097·0	·03540318	716·00	711·50	4·50	26·67	·17	100·6
72	19073·1	·03909805	717·01	745·72	−28·72	27·31	−1·05	96·1
73	17975·3	·04315694	782·64	775·76	6·88	27·85	·25	100·9
74	16750·7	·04761489	797·87	797·58	·29	28·24	·01	100·0
75	14496·5	·05251033	801·39	761·22	40·17	27·59	1·46	105·3
76	12701·9	·05788538	762·50	735·25	27·25	27·12	1·00	103·7
77	10907·8	·06378620	742·57	695·77	46·80	26·38	1·77	106·7
78	9356·5	·07026341	633·77	657·42	−23·64	25·64	−·92	96·4
79	8479·5	·07737250	702·05	656·08	45·98	25·61	1·80	107·0
80	7171·3	·08517430	676·22	610·81	65·41	24·71	2·65	110·7
81	5413·2	·09373550	550·31	507·41	42·91	22·53	1·90	108·5
82	5217·7	·10312921	544·29	538·10	6·19	23·20	·27	101·2
83	4641·5	·11343557	565·41	526·52	38·90	22·95	1·70	107·4
84	3622·5	·12474245	466·19	451·88	14·31	21·26	·67	103·2
85	2981·0	·13714617	386·52	408·84	−22·31	20·22	−1·10	94·5
86	2578·3	·15075231	376·30	388·69	−12·39	19·72	−·63	96·8
87	2303·3	·16567664	356·69	381·61	−24·92	19·53	−1·28	93·5
88	1954·5	·18204606	322·95	355·81	−32·86	18·86	−1·74	90·8
89	1481·1	·19999969	290·91	296·21	−5·30	17·21	−·31	98·2
90	1192·6	·21969004	230·88	262·01	−31·13	16·19	−1·92	88·1
T.	10823308·0		52379·83	52379·83	·00			100·0

$\chi^2 = \Sigma z_x^2 = 109{\cdot}01$

added parameters. None meets all the desired criteria, but GM(0, 8) ($L_1 = -20{\cdot}1$) and GM(3, 6) ($L_1 = -17{\cdot}0$) meet many of them. The GM(3, 6) formula has the highest value of $p(\chi^2)$ of any formula, but it must be remembered that the value of χ^2 is affected by our adjustment for duplicates, and it is by no means certain that our adjustment (being based only on the deaths) is correct.

Table 17.10. *Male assured lives, permanent, U.K., Duration 5 +. Details of graduation Aa1. Function: μ_x, GM(3, 6), criterion: maximum likelihood. Ages grouped so that each $E_x \geqslant 5$. Age range 10–90 only*

Age	R'_x	μ_x	A'_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	100A/E
10	117·0	·00252398	·00	·30	− ·30			
11	33·5	·00232998	·00	·08	− 08			
12	49·7	·00214603	·00	·11	− ·11			
13	56·8	·00197214	·00	·11	− ·11			
14	79·7	·00180831	1·00	·14	·86			
15	238·0	·00165454	1·00	·39	·61			
16	538·7	·00151082	1·00	·81	·19			
17	860·5	·00137716	·00	1·19	− 1·19			
18	1137·5	·00125356	2·00	1·43	·57			
19	1433·5	·00114002	·00	1·63	− 1·63			
10–19	4544·9		5·00	6·19	− 1·19	2·49	− ·48	80·8
20	2330·0	·00103653	3·00	2·42	·58			
21	5478·7	·00094309	2·40	5·17	− 2·77			
20–21	7808·7		5·40	7·58	− 2·19	2·75	− ·79	71·2
22	24755·4	·00085972	23·97	21·28	2·68	4·61	·58	112·6
23	45526·6	·00078640	41·07	35·80	5·27	5·98	·88	114·7
24	74643·0	·00072314	52·00	53·98	− 1·98	7·35	− ·27	96·3
25	85112·7	·00066993	61·86	57·02	4·84	7·55	·64	108·5
26	96296·3	·00062679	60·45	60·36	·09	7·77	·01	100·1
27	133767·8	·00059370	77·87	79·42	− 1·55	8·91	− ·17	98·1
28	159323·6	·00057066	77·95	90·92	− 12·97	9·54	− 1·36	85·7
29	189169·1	·00055768	116·79	105·50	11·30	10·27	1·10	110·7
30	238694·0	·00055476	126·98	132·42	− 5·43	11·51	− ·47	95·9
31	295929·1	·00056190	151·22	166·28	− 15·06	12·90	− 1·17	90·9
32	284384·9	·00057910	160·39	164·69	− 4·30	12·83	− ·33	97·4
33	312853·4	·00060635	211·25	189·70	21·55	13·77	1·56	111·4
34	399268·0	·00064366	254·07	256·99	− 2·92	16·03	− ·18	98·9
35	363515·5	·00069106	278·06	251·21	26·85	15·85	1·69	110·7
36	367216·9	·00074860	258·44	274·90	− 16·46	16·58	− ·99	94·0
37	376698·0	·00081643	301·35	307·55	− 6·20	17·54	− ·35	98·0
38	302206·2	·00089490	295·05	270·44	24·61	16·45	1·50	109·1
39	317344·8	·00098471	285·71	312·49	− 26·78	17·68	− 1·51	91·4
40	328307·0	·00108717	337·34	356·93	− 19·58	18·89	− 1·04	94·5
41	300565·3	·00120439	377·33	362·00	15·33	19·03	·81	104·2
42	292776·8	·00133947	399·44	392·17	7·28	19·80	·37	101·9
43	299940·9	·00149656	430·73	448·88	− 18·15	21·19	− ·86	96·0
44	314046·6	·00168069	549·12	527·82	21·31	22·97	·93	104·0
45	316713·0	·00189740	607·78	600·93	6·85	24·51	·28	101·1
46	278217·8	·00215230	612·83	598·81	14·03	24·47	·57	102·3
47	303899·8	·00245050	703·55	744·71	− 41·16	27·29	− 1·51	94·5
48	292654·6	·00279626	762·64	818·34	− 55·70	28·61	− 1·95	93·2
49	281024·7	·00319271	945·56	897·23	48·32	29·95	1·61	105·4
50	274140·7	·00364188	1026·37	998·39	27·99	31·60	·89	102·8
51	281170·4	·00414487	1204·02	1165·41	38·61	34·14	1·13	103·3

Table 17.10 (*cont.*)

Age	R'_x	μ_x	A'_x	E_x	Dev_x	$(V_x)^{1/2}$	z_x	$100A/E$
52	265230·1	·00470226	1245·56	1247·18	−1·62	35·32	−·05	99·9
53	280678·5	·00531452	1510·84	1491·67	19·17	38·62	·50	101·3
54	275102·4	·00598253	1609·70	1645·81	−36·11	40·57	−·89	97·8
55	259981·3	·00670799	1713·17	1743·95	−30·78	41·76	−·74	98·2
56	232766·6	·00749381	1734·64	1744·31	−9·67	41·76	−·23	99·4
57	231145·5	·00834442	1883·33	1928·78	−45·44	43·92	−1·03	97·6
58	231889·2	·00926605	2106·51	2148·70	−42·19	46·35	−·91	98·0
59	220517·6	·01026692	2333·53	2264·04	69·49	47·58	1·46	103·1
60	206083·5	·01135740	2290·32	2340·57	−50·25	48·38	−1·04	97·9
61	171438·7	·01255022	2155·48	2151·59	3·89	46·39	·08	100·2
62	148951·6	·01386059	2136·13	2064·56	71·57	45·44	1·58	103·5
63	138082·9	·01530638	2236·30	2113·55	122·75	45·97	2·67	105·8
64	117520·6	·01690828	1971·24	1987·07	−15·83	44·58	−·36	99·2
65	73385·6	·01868996	1369·34	1371·57	−2·23	37·03	−·06	99·8
66	45338·4	·02067812	892·00	937·51	−45·51	30·62	−1·49	95·1
67	33737·4	·02290260	729·50	772·67	−43·18	27·80	−1·55	94·4
68	27694·1	·02539627	697·97	703·33	−5·35	26·52	−·20	99·2
69	27855·9	·02819485	765·91	785·39	−19·48	28·02	−·70	97·5
70	24996·2	·03133647	774·24	783·29	−9·05	27·99	−·32	98·8
71	20097·0	·03486105	716·00	700·60	15·40	26·47	·58	102·2
72	19073·1	·03880935	717·01	740·22	−23·21	27·21	−·85	96·9
73	17975·3	·04322171	782·64	776·93	5·71	27·87	·20	100·7
74	16750·7	·04813643	797·87	806·32	−8·45	28·40	−·30	99·0
75	14496·5	·05358780	801·39	776·84	24·55	27·87	·88	103·2
76	12701·9	·05960400	762·50	757·08	5·42	27·52	·20	100·7
77	10907·8	·06620464	742·57	722·14	20·42	26·87	·76	102·8
78	9356·5	·07339851	633·77	686·75	−52·98	26·21	−2·02	92·3
79	8479·5	·08118152	702·05	688·37	13·68	26·24	·52	102·0
80	7171·3	·08953512	676·22	642·09	34·14	25·34	1·35	105·3
81	5413·2	·09842568	550·31	532·80	17·52	23·08	·76	103·3
82	5217·7	·10780499	544·29	562·49	−18·21	23·72	−·77	96·8
83	4641·5	·11761219	565·41	545·90	19·51	23·36	·84	103·6
84	3622·5	·12777739	466·19	462·87	3·32	21·51	·15	100·7
85	2981·0	·13822704	386·52	412·06	−25·53	20·30	−1·26	93·8
86	2578·3	·14889092	376·30	383·89	−7·59	19·59	−·39	98·0
87	2303·3	·15971085	356·69	367·87	−11·18	19·18	−·58	97·0
88	1954·5	·17065051	322·95	333·54	−10·59	18·26	−·58	96·8
89	1481·1	·18170652	290·91	269·12	21·79	16·40	1·33	108·1
90	1192·6	·19292030	230·88	230·09	·80	15·17	·05	100·3
T.	10823308·0		52379·83	52379·83	·00			100·0

$\chi^2 = \Sigma z_x^2 = 69{\cdot}40$

The value of L_1 for GM(3, 6) is 20·5 higher than that for GM(2, 2). The likelihood ratio test (see § 10.2) compares twice the difference between the values of L_1 with $\chi^2(5)$, since there are five extra parameters in the GM(3, 6) formula. There is a very significant improvement. Strictly the likelihood ratio test should not be used to compare the GM(0, 8) with either the GM(0, 2) nor the GM(3,

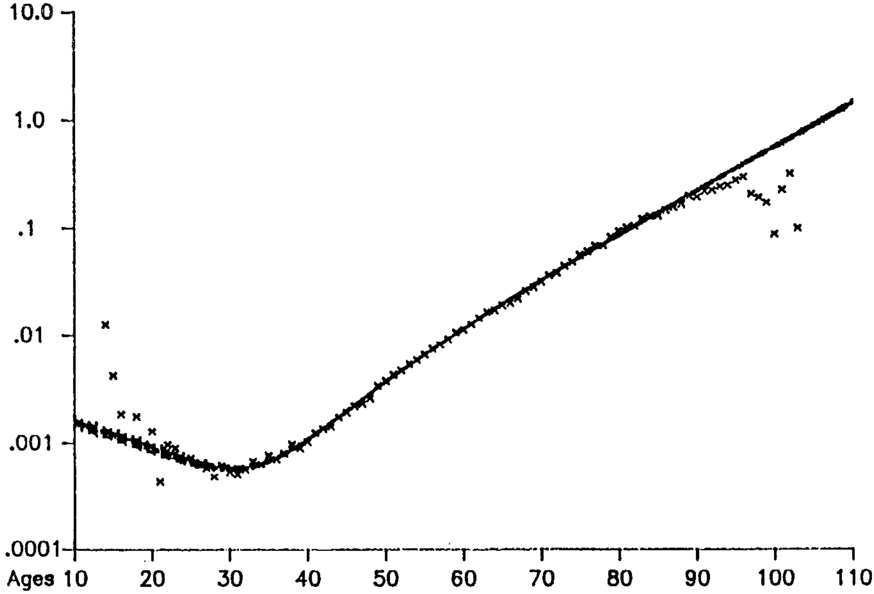

Figure 17.6. Male assured lives Durations 5+. Sheaf for μ_x = GM(2, 2).

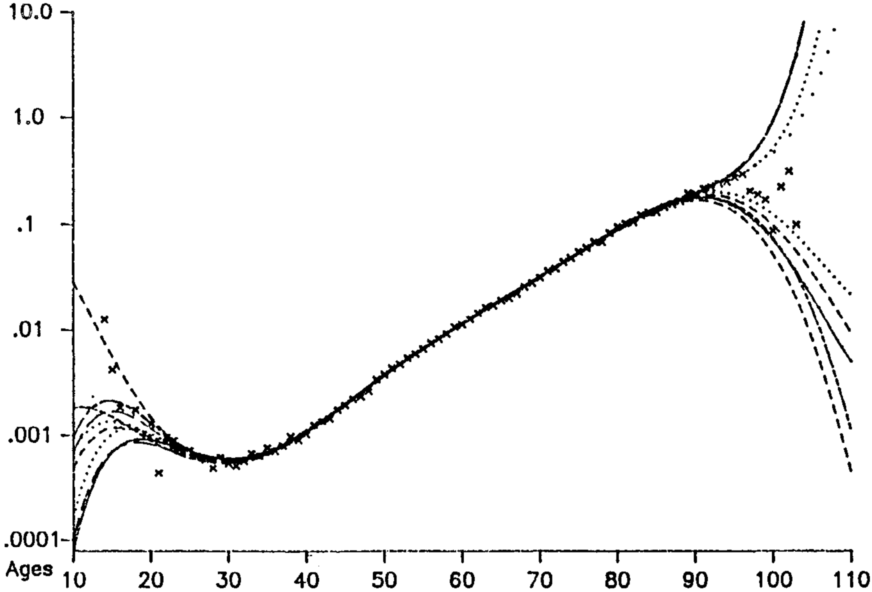

Figure 17.7. Male assured lives Durations 5+. Sheaf for μ_x = GM(3, 6).70.

6) formulae, since it is not strictly of a higher or lower order in both dimensions (r and s) than either of them.

Although the final s-parameter of the GM(3, 6) formula is only marginally significant, the shape of the formula makes it worth considering. Statistics for the GM(2, 2), GM(0, 8) and GM(3, 6) graduations are shown in Table 17.8. Details of the GM(2, 2) and GM(3, 6) graduations are shown in Tables 17.9 and 17.10. The calculated values of μ_x according to the GM(2, 2) formula are shown in Figure 17.5, and the corresponding sheaf is shown in Figure 17.6. The sheaf for GM(3, 6) is shown in Figure 17.7.

The methods used for graduating duration 5+ are equally appropriate for graduating duration 2+, and the results are very similar. This is discussed in the accompanying CMI Report.

The 'distance' between the parameters of the GM(2, 2) graduation for duration 5+ can be compared with those for the select durations, as was done when comparing the select durations among themselves. The resulting values of D are all very large, so the parameters for 5+ are clearly different from those for any of the select durations.

ACKNOWLEDGEMENT

We are grateful to several colleagues for instructive and stimulating discussions on various topics relating to this paper. In particular we thank Mr. H. A. R. Barnett, Secretary of the CMI Bureau, whose considerable knowledge was freely made available to us, and Mr. P. A. Leandro, who assisted the Graduation Working Party.

We are grateful to our colleagues on the CMI Committee for permission to use CMI data to illustrate our results. We thank also Dr. I. D. Currie and Dr. S. Zachary, of Heriot-Watt University, for guidance on certain statistical matters.

Finally, we must record our appreciation of the secretarial skills of Mrs. W. Hughes and Mrs. M. Flight, who so efficiently produced the typed manuscript.

REFERENCES

AKAIKE, H. (1978). On the likelihood of a Time-series Model. *The Statistician*, **27**, 217.

AKAIKE, H. (1985). 'Prediction and Entropy' in A Celebration of Statistics, edited by ATKINSON, C. & FIENBERG, S. E., Springer Verlag, New York.

BARNETT, H. A. R. (1951). Graduation Tests and Experiments, *J.I.A.*, **77**, 15.

BARNETT, H. A. R. (1985). Criteria of Smoothness, *J.I.A.*, **112**, 331.

BATTEN, R. W. (1978). Mortality Table Construction. Prentice-Hall, Englewood Cliffs, N.J.

BEARD, R. E. & PERKS, W. (1949). The Relation between the Distribution of Sickness and the Effect of Duplicates on the Distribution of Deaths. *J.I.A.*, **75**, 75.

BENJAMIN, B. & POLLARD, J.H. (1980). The Analysis of Mortality and Other Actuarial Statistics, Second Edition. Heinemann, London.

CMI Committee (1957). Continuous Investigation into the Mortality of Assured Lives, Memorandum on a Special Inquiry into the Distribution of Duplicate Policies. *J.I.A.*, **83**, 34 and *T.F.A.*, **24**, 94.

CMI Committee (1974). Considerations Affecting the Preparation of Standard Tables of Mortality, *J.I.A.*, **101**, 133 and *T.F.A.*, **34**, 135.

CMI Committee (1976). The Graduation of Pensioners' and of Annuitants' Mortality Experience 1967–70, *CMIR*, **2**, 57.

CMI Committee (1986). An investigation into the distribution of policies per life assured in the cause of death investigation data, *CMIR*, **8**, 49.

CMI Committee (1986). The C.M.I. Bureau: A Note on the History of the Computerization of the Work of the Bureau and the Development of Improved Services to Contributing Offices (Appendix: The Kolmogorov–Smirnov Test), *CMIR*, **8**, 59.

CONTE, S. D. & de BOOR, C. (1980), *Elementary Numerical Analysis*, Third Edition. McGraw-Hill Kogakusha, Tokyo.

COX, D. R. & OAKES, D. (1984). Analysis of Survival Data. Chapman and Hall, London.

DAW, R. H. (1946). On the Validity of Statistical Tests of the Graduation of a Mortality Table. *J.I.A.*, **72**, 174.

DAW, R. H. (1951). Duplicate Policies in Mortality Data. *J.I.A.*, **77**, 261.

De BOOR, C. (1978). *A Practical Guide to Splines*. Springer-Verlag, New York.

DURBIN, J. (1973). Distribution Theory for Tests Based on the Sample Distribution Function. Society for Industrial and Applied Mathematics, Philadelphia, PA.

ELANDT-JOHNSON, R. C. & JOHNSON, N. L. (1980). Survival Models and Data Analysis. John Wiley, New York.

EDWARDS, A. W. F. (1972). Likelihood. Cambridge University Press.

FISZ, M. (1963). Probability Theory and Mathematical Statistics. John Wiley, New York.

FORFAR, D. O. & SMITH D. M. (1987). The changing shape of English Life Tables *T.F.A.*, **40**, 98.

GREVILLE, T. N. E. (1978). Estimation of the Rate of Mortality in the Presence of In-and-out Movement, *ARCH*, 1978.2, 41.

HELIGMAN, L. & POLLARD, J. H. (1980). The Age Pattern of Mortality, *J.I.A.*, **107**, 49.

HOEM, J. (1980). Exposed-to-risk Considerations based on the Balducci assumption and other assumptions in the analysis of mortality, *ARCH*, 1980. 1, 47.

HOEM J. (1984). A Flaw in Actuarial Exposed-to-Risk Theory. *Scandinavian Actuarial Journal*, **1984**, 187.

HOGG, R. V. & KLUGMAN, S. A. (1984). Loss Distributions. John Wiley, New York.

KALBFLEISCH, J. D. & PRENTICE, R. L. (1980). The Statistical Analysis of Failure Time Data. John Wiley, New York.

KENDALL, M. G. & STUART, A. (1979). The Advanced Theory of Statistics, Volume 2, Fourth Edition. Charles Griffin, London.

LARSON, H. J. (1982). Introduction to Probability Theory and Statistical Inference, Third Edition. John Wiley, New York.

LONDON, D. (1985). Graduation: The Revision of Estimates. Actex Publications, Winsted and Abington, Conn.

McCUTCHEON, J. J. (1971). Some remarks on the basic mortality functions, *T.F.A.*, **32**, 395.

McCUTCHEON, J. J. (1977). Some Elementary Life Table Approximations, *T.F.A.*, **35**, 297.

McCUTCHEON, J. J. (1981). Some Remarks on Splines, *T.F.A.*, **37**, 421.

McCUTCHEON, J. J. (1982). Graduation of the Experience of Female Assured Lives 1975–78, *T.F.A.*, **38**, 193.

McCUTCHEON, J. J. (1983). On estimating the Force of Mortality. *T.F.A.*, **38**, 407.

McCUTCHEON, J. J. (1984). Spline Graduation with Variable Knots, *Proceedings 22nd I.C.A.*, **4**, 47.

McCUTCHEON, J. J. (1987). Experiments in graduating the data for the English Life Tables No. 14, *T.F.A.* **40**, 135.

MANN, N. R., SCHAFER, R. E. & SINGPURWALLA, N. D. (1974). Methods for Statistical Analysis of Reliability and Life Data. John Wiley, New York.

Maturity Guarantees Working Party (1980). Report, *J.I.A.*, **107**, 101.

NELDER, J. A. & MEAD, R. (1965). A simplex method for function minimization. *Computer Journal*, **7**, 308.

PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTERLING, W. T. (1986). Numerical Recipes. Cambridge University Press.

RAO, C. R. (1973). Linear Statistical Inference and its Applications, Second Edition. John Wiley, New York.

RAO, S. S. (1984). Optimization Theory and Applications, Second Edition. Wiley Eastern Ltd, New Delhi.

ROBERTS, L. A. (1986). Bias in decremental rate estimates, *OARD* (**39**), Institute of Actuaries.

RUBINSTEIN, R. Y. (1986). Simulation and the Monte Carlo Method. John Wiley, New York.

SEAL, H. (1945). Tests of a mortality table graduation. *J.I.A.*, **71**, 3.

SCOTT, W. F. (1982). Some Applications of the Poisson Distribution in Mortality Studies, *T.F.A.*, **38**, 255.

SCOTT, W. F. (1986). On the calculation of the Exposed-to-Risk in the CMI investigations. Unpublished note.

SVERDRUP, E. (1965). Estimates and Test Procedure in Connection with Stochastic Models for Deaths, Recoveries and Transfers between different States of Health, *Skandinavisk Aktuarietidskrift*, **48**, 184.

Van der WAERDEN, B. L. (1969). Mathematical Statistics. George Allan and Unwin, London.

WATERS, H. R. & WILKIE, A. D. (1987). A short note on the construction of life tables and multiple decrement tables. *J.I.A.*, **114**, 569.

APPENDIX 1

Derivation of the information matrix $H(\boldsymbol{\alpha})$ for each of the models described in § 6.

A1 Remark on notation

For notational simplification throughout this appendix we write A, R, r, q, μ, and m for A_x, R_x, r_x, $q_{x+b}(\boldsymbol{\alpha})$, $\mu_{x+b+1/2}(\boldsymbol{\alpha})$, and $m_{x+b}(\boldsymbol{\alpha})$ respectively. We write q^i and q^{ij} to denote $\partial q/\partial\alpha_i$ and $\partial^2 q/\partial\alpha_i\partial\alpha_j$ respectively, while μ^i, μ^{ij}, m^i and m^{ij} are similarly defined.

A2 q-type rates (initial exposures - see § 6.2)

In this situation $L^*(\boldsymbol{\alpha})$ is defined in terms of the values of q at each age. (See equations (6.2.4), (6.2.15), and (6.2.16.)) Let s be a typical term in the sum which defines $L^*(\boldsymbol{\alpha})$.

Note that

$$\frac{\partial s}{\partial\alpha_i} = \frac{\partial s}{\partial q}q^i \tag{1}$$

and

$$\frac{\partial^2 s}{\partial\alpha_i\partial\alpha_j} = \frac{\partial s}{\partial q}q^{ij} + \frac{\partial^2 s}{\partial q^2}q^iq^j \tag{2}$$

Thus (see equation (8.2.1))

$$H^*_{ij}(\boldsymbol{\alpha}) = -\Sigma\left(\frac{\partial s}{\partial q}q^{ij} + \frac{\partial^2 s}{\partial q^2}q^iq^j\right) \tag{3}$$

the summation being over the range of ages appropriate to the definition of $L^*(\boldsymbol{\alpha})$.

(i) $L^*(\boldsymbol{\alpha}) = L_1(\boldsymbol{\alpha})$ (binomial model with no duplicates)

In this case (see equation (6.2.4))

$$s = A\log q + (R - A)\log(1 - q) \tag{4}$$

Now

$$\frac{\partial s}{\partial q} = \frac{A}{q} - \frac{(R - A)}{(1 - q)} \tag{5}$$

and

$$\frac{\partial^2 s}{\partial q^2} = -\frac{A}{q^2} - \frac{(R - A)}{(1 - q)^2} \tag{6}$$

Combining equations (3), (5), and (6), we obtain

$$H^*_{ij}(\boldsymbol{\alpha}) = \sum\left\{\left(\frac{R - A}{1 - q} - \frac{A}{q}\right)q^{ij} + \left(\frac{A}{q^2} + \frac{R - A}{(1 - q)^2}\right)q^iq^j\right\} \tag{7}$$

Since at each age $E[A] = Rq$, on taking expected values we find

$$H_{ij}(\alpha) = \sum \left\{ \left(\frac{R - Rq}{1 - q} - \frac{Rq}{q} \right) q^{ij} + \left(\frac{Rq}{q^2} + \frac{R - Rq}{(1 - q)^2} \right) q^i q^j \right\}$$

$$= \sum \frac{R}{q(1 - q)} q^i q^j \tag{8}$$

It should be noted that this last expression does *not* require the calculation of the second order partial derivative q^{ij}.

(ii) $L^*(\alpha) = L_2(\alpha)$ (normal approximation and allowance for duplicates)
In this case (see equation (6.2.15))

$$s = -\frac{1}{2} \left\{ \log q + \log (1 - q) + \frac{(A - Rq)^2}{rRq(1 - q)} \right\} \tag{9}$$

Now

$$\frac{\partial s}{\partial q} = -\frac{1}{2} \left\{ \frac{1}{q} - \frac{1}{1 - q} + \frac{-2R(A - Rq)q(1 - q) - (A - Rq)^2(1 - 2q)}{rRq^2(1 - q)^2} \right\}$$

$$= -\frac{1}{2} \left\{ \frac{1}{q} - \frac{1}{1 - q} + \frac{(R - A)^2 q^2 - A^2(1 - q)^2}{rRq^2(1 - q)^2} \right\}$$

$$= -\frac{1}{2} \left\{ \frac{1}{q} - \frac{1}{1 - q} + \frac{(R - A)^2}{rR(1 - q)^2} - \frac{A^2}{rRq^2} \right\} \tag{10}$$

and

$$\frac{\partial^2 s}{\partial q^2} = -\frac{1}{2} \left\{ -\frac{1}{q^2} - \frac{1}{(1 - q)^2} + \frac{2(R - A)^2}{rR(1 - q)^3} + \frac{2A^2}{rRq^3} \right)$$

$$= \frac{1}{2} \left(\frac{1}{q^2} + \frac{1}{(1 - q)^2} \right) - \left(\frac{(R - A)^2}{rR(1 - q)^3} + \frac{A^2}{rRq^3} \right) \tag{11}$$

Combining equation (3), (10), and (11), we obtain

$$H^*_{ij}(\alpha) = \sum \left\{ \frac{1}{2} \left[\frac{1}{q} - \frac{1}{1 - q} + \frac{(R - A)^2}{rR(1 - q)^2} - \frac{A^2}{rRq^2} \right] q^{ij} + \left[-\frac{1}{2} \left(\frac{1}{q^2} + \frac{1}{(1 - q)^2} \right) + \frac{(R - A)^2}{rR(1 - q)^3} + \frac{A^2}{rRq^3} \right] q^i q^j \right\} \tag{12}$$

Now

$$E[A] = Rq$$

and

$$\mathrm{Var}\,[A] = rRq(1 - q)$$

Hence

$$E[A^2] = R^2q^2 + rRq(1 - q) \tag{13}$$

and

$$\begin{aligned} E[(R - A)^2 &= R^2 - 2RE[A] + E[A^2] \\ &= R^2 - 2R^2q + R^2q^2 + rRq(1 - q) \\ &= R^2(1 - q)^2 + rRq(1 - q) \end{aligned} \tag{14}$$

To determine the information matrix, $H(\alpha)$, we take expected values of equation (12). Using equations (13) and (14) we find by simple algebra that the expected value of the coefficient of q^{ij} in equation (12) is zero, while the coefficient of q^iq^j has expected value

$$\begin{aligned} & -\frac{1}{2}\left(\frac{1}{q^2} + \frac{1}{(1 - q)^2}\right) + \frac{R^2(1 - q)^2 + rRq(1 - q)}{rR(1 - q)^3} \\ & + \frac{R^2q^2 + rRq(1 - q)}{rRq^3} \\ & = \frac{(1 - 2q)^2}{2q^2(1 - q)^2} + \frac{R}{rq(1 - q)} \end{aligned} \tag{15}$$

Hence, taking expected values of equation (12), we obtain

$$H_{ij}(\alpha) = \sum\left\{\frac{(1 - 2q)^2}{2q^2(1 - q)^2} + \frac{R}{rq(1 - q)}\right\}q^iq^j \tag{16}$$

Again it should be noted that this last equation does *not* involve any second order partial derivatives.

(iii) $L^*(\alpha) = L_3(\alpha)$ (further approximation and allowance for duplicates) In this case (see equation (6.2.16)

$$s = -\frac{1}{2}\frac{(A - Rq)^2}{rRq(1 - q)} \tag{17}$$

This is simply the final term in equation (9) above. Hence

$$\frac{\partial s}{\partial q} = -\frac{1}{2}\left\{\frac{(R - A)^2}{rR(1 - q)^2} - \frac{A^2}{rRq^2}\right\} \tag{18}$$

and

$$\frac{\partial^2 s}{\partial q^2} = -\left(\frac{(R-A)^2}{rR(1-q)^3} + \frac{A^2}{rRq^3}\right) \tag{19}$$

Equations (3), (18), and (19) above imply that

$$H^*_{ij}(\boldsymbol{\alpha}) = \sum\left\{\frac{1}{2}\left[\frac{(R-A)^2}{rR(1-q)^2} - \frac{A^2}{rRq^2}\right]q^{ij} + \left[\frac{(R-A)^2}{rR(1-q)^3} + \frac{A^2}{rRq^3}\right]q^i q^j\right\} \tag{20}$$

Taking expected values, we obtain

$$H_{ij}(\boldsymbol{\alpha}) = \sum\left\{-\frac{1-2q}{2q(1-q)}q^{ij} + \left[\frac{1-3q+3q^2}{q^2(1-q)^2} + \frac{R}{rq(1-q)}\right]q^i q^j\right\} \tag{21}$$

Note that, in contrast to equations (8) and (16) above, this last expression *does* involve the second order partial derivative q^{ij}

A3 μ-type rates (cental exposures - see § 6.3)

In this model $L^*(\boldsymbol{\alpha})$ is defined in terms of μ and (cf equation (3) above)

$$H^*_{ij}(\boldsymbol{\alpha}) = -\sum\left(\frac{\partial s}{\partial \mu}\mu^{ij} + \frac{\partial^2 s}{\partial \mu^2}\mu^i\mu^j\right) \tag{22}$$

where s is a typical term in the sum which defines $L^*(\boldsymbol{\alpha})$.

(i) $L^*(\boldsymbol{\alpha}) = L_1(\boldsymbol{\alpha})$ (Poisson model with no duplicates)

In this case (see equation (6.3.5)

$$s = -R\mu + A\log\mu \tag{23}$$

Clearly

$$\frac{\partial s}{\partial \mu} = -R + \frac{A}{\mu} \tag{24}$$

and

$$\frac{\partial^2 s}{\partial \mu^2} = -\frac{A}{\mu^2} \tag{25}$$

Combining these last two equation with equation (22) above, we obtain

$$H^*_{ij}(\boldsymbol{\alpha}) = \sum\left\{\left[R - \frac{A}{\mu}\right]\mu^{ij} + \frac{A}{\mu^2}\mu^i\mu^j\right\} \tag{26}$$

Since, at each age, $E[A] = R\mu$, on taking expected values we have

$$H_{ij}(\boldsymbol{\alpha}) = \sum \frac{R}{\mu} \mu^i \mu^j \tag{27}$$

(ii) $L^*(\boldsymbol{\alpha}) = L_2(\boldsymbol{\alpha})$ (normal approximation and allowance for duplicates)
In this case (see equation (6.3.16))

$$s = -\frac{1}{2}\left\{\log \mu + \frac{(A - R\mu)^2}{rR\mu}\right\} \tag{28}$$

Now

$$\frac{\partial s}{\partial \mu} = -\frac{1}{2}\left\{\frac{1}{\mu} + \frac{-2R\mu(A - R\mu) - (A - R\mu)^2}{rR\mu^2}\right\}$$

$$= -\frac{1}{2}\left\{\frac{1}{\mu} - \frac{A^2}{rR\mu^2} + \frac{R}{r}\right\} \tag{29}$$

and

$$\frac{\partial^2 s}{\partial \mu^2} = -\left\{-\frac{1}{2\mu^2} + \frac{A^2}{rR\mu^3}\right\} \tag{30}$$

Combining these last two equations with equation (22) above, we obtain

$$H^*_{ij}(\boldsymbol{\alpha}) = \sum\left\{\frac{1}{2}\left(\frac{1}{\mu} - \frac{A^2}{rR\mu^2} + \frac{R}{r}\right)\mu^{ij} + \left(-\frac{1}{2\mu^2} + \frac{A^2}{rR\mu^3}\right)\mu^i\mu^j\right\} \tag{31}$$

Since at each age $E[A] = R\mu$ and $\text{Var}[A] = rR\mu$,

$$E[A^2] = R^2\mu^2 + rR\mu$$

Accordingly, on taking expected values in equation (31) we find

$$H_{ij}(\boldsymbol{\alpha}) = \sum\left\{\frac{1}{2\mu^2} + \frac{R}{r\mu}\right\}\mu^i\mu^j \tag{32}$$

(iii) $L^*(\boldsymbol{\alpha}) = L_3(\boldsymbol{\alpha})$ (further approximation and allowance for duplicates)
In this case (see equation (6.3.17)

$$s = -\frac{1}{2}\frac{(A - R\mu)^2}{rR\mu} \tag{33}$$

This is simply the second term in equation (28) above and (cf. equation (29) and (30)) therefore

$$\frac{\partial s}{\partial \mu} = -\frac{1}{2}\left\{-\frac{A^2}{rR\mu^2} + \frac{R}{r}\right\} \tag{34}$$

and

$$\frac{\partial^2 s}{\partial \mu^2} = -\frac{A^2}{rR\mu^3} \tag{35}$$

Combining these last two equations with equation (22) above, we obtain

$$H^*_{ij}(\boldsymbol{\alpha}) = \sum \left\{ \frac{1}{2}\left(-\frac{A^2}{rR\mu^2} + \frac{R}{r} \right) \mu^{ij} + \frac{A^2}{rR\mu^3}\mu^i \mu^j \right\} \tag{36}$$

Taking expected values, we get

$$H_{ij}(\boldsymbol{\alpha}) = \sum \left\{ -\frac{1}{2\mu}\mu^{ij} + \left(\frac{1}{\mu^2} + \frac{R}{r\mu} \right) \mu^i \mu^j \right\} \tag{37}$$

A4 m-type rates (central exposures - see § 6.4)

Here $L^*(\boldsymbol{\alpha})$ is defined in terms of the values of m at each age. We repeat the discussion of § A3 above, simply replacing μ by m throughout. Thus

(i) $L^*(\boldsymbol{\alpha}) = L_1(\alpha)$ (Poisson model with no duplicates)

$$H_{ij}(\boldsymbol{\alpha}) = \sum \frac{R}{m} m^i m^j \tag{38}$$

(ii) $L^*(\boldsymbol{\alpha}) = L_2(\boldsymbol{\alpha})$ (normal approximation and allowance for duplicates)

$$H_{ij}(\boldsymbol{\alpha}) = \sum \left\{ \frac{1}{2m^2} + \frac{R}{rm} \right\} m^i m^j \tag{39}$$

(iii) $L^*(\boldsymbol{\alpha}) = L_3(\boldsymbol{\alpha})$ (further approximation and allowance for duplicates)

$$H_{ij}(\boldsymbol{\alpha}) = \sum \left\{ -\frac{1}{2m} m^{ij} + \left(\frac{1}{m^2} + \frac{R}{rm} \right) m^i m^j \right\} \tag{40}$$

A5 q-type rates (central exposures - see § 6.5)

In this case the log-likelihood $L^*(\boldsymbol{\alpha})$ is defined in terms of the values of q at each age. In evaluating the partial derivatives of $L^*(\boldsymbol{\alpha})$ we find it convenient to make use of the results in § A3 above. Note that $L^*(\boldsymbol{\alpha})$ may be defined in terms of the values of μ at each age, where

$$\mu = -\log(1 - q) \tag{41}$$

Hence

$$\mu^i = \frac{1}{1-q} q^i \tag{42}$$

and

$$\mu^{ij} = \frac{1}{1-q} q^{ij} + \frac{1}{(1-q)^2} q^i q^j \tag{43}$$

Let s be a typical term in the sum which defines $L^*(\boldsymbol{\alpha})$, s being expressed in terms of $\mu = -\log(1-q)$. Then (cf. equation (2) above)

$$\frac{\partial^2 s}{\partial\alpha_i \partial\alpha_j} = \frac{\partial s}{\partial\mu}\mu^{ij} + \frac{\partial^2 s}{\partial\mu^2}\mu^i\mu^j$$

Combining this last equation with equations (42) and (43), we have

$$\begin{aligned}\frac{\partial^2 s}{\partial\alpha_i \partial\alpha_j} &= \frac{\partial s}{\partial\mu}\left[\frac{1}{1-q}q^{ij} + \frac{1}{(1-q)^2}q^iq^j\right] + \frac{\partial^2 s}{\partial\mu^2}\left[\frac{1}{(1-q)^2}q^iq^j\right] \\ &= \frac{1}{1-q}\frac{\partial s}{\partial\mu}q^{ij} + \frac{1}{(1-q)^2}\left(\frac{\partial s}{\partial\mu} + \frac{\partial^2 s}{\partial\mu^2}\right)q^iq^j \qquad (44)\end{aligned}$$

Thus (see equation (8.2.1))

$$H^*_{ij}(\boldsymbol{\alpha}) = -\sum\left\{\frac{1}{1-q}\frac{\partial s}{\partial\mu}q^{ij} + \frac{1}{(1-q)^2}\left(\frac{\partial s}{\partial\mu} + \frac{\partial^2 s}{\partial\mu^2}\right)q^iq^j\right\} \qquad (45)$$

In this last equation, for each model the partial derivatives are those determined in § A3 above, but in their evaluation we must substitute $-\log(1-q)$ for μ. These remarks lead immediately to the following results.

(i) $L^*(\boldsymbol{\alpha}) = L_1(\boldsymbol{\alpha})$ (Poisson model with no duplicates).
(See equation (6.5.2).)
It follows from equations (24) and (25) that

$$\frac{\partial s}{\partial\mu} = -R + \frac{A}{\mu} = -R - \frac{A}{\log(1-q)}$$

and

$$\frac{\partial^2 s}{\partial\mu^2} = -\frac{A}{\mu^2} = -\frac{A}{[\log(1-q)]^2}$$

Combining these last two equations with equation (45), we obtain

$$\begin{aligned}H^*_{ij}(\boldsymbol{\alpha}) = \sum\Bigg\{&\frac{1}{1-q}\left(R + \frac{A}{\log(1-q)}\right)q^{ij} \\ &+ \frac{1}{(1-q)^2}\Bigg(R + \frac{A}{\log(1-q)} \\ &\qquad + \frac{A}{[\log(1-q)]^2}\Bigg)q^iq^j\Bigg\} \qquad (46)\end{aligned}$$

Since $E[A] = R\mu = -R\log(1-q)$, on taking expected values we obtain

$$H_{ij}(\boldsymbol{\alpha}) = -\sum\frac{R}{(1-q)^2[\log(1-q)]}q^iq^j \qquad (47)$$

(ii) $L^*(\boldsymbol{\alpha}) = L_2(\boldsymbol{\alpha})$ (normal approximation and allowance for duplicates). (See equation (6.5.4).)

In this situation equations (29) and (30) imply that

$$\begin{aligned}\frac{\partial s}{\partial \mu} &= -\frac{1}{2}\left\{\frac{1}{\mu} - \frac{A^2}{rR\mu^2} + \frac{R}{r}\right\} \\ &= -\frac{1}{2}\left\{-\frac{1}{\log(1-q)} - \frac{A^2}{rR[\log(1-q)]^2} + \frac{R}{r}\right\}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2 s}{\partial \mu^2} &= -\left\{-\frac{1}{2\mu^2} + \frac{A^2}{rR\mu^3}\right\} \\ &= \frac{1}{2[\log(1-q)]^2} + \frac{A^2}{rR[\log(1-q)]^3}\end{aligned}$$

Substituting these partial derivatives in equation (45), we obtain

$$\begin{aligned}H^*_{ij}(\boldsymbol{\alpha}) = \sum \Bigg\{&\frac{1}{2(1-q)}\left(-\frac{1}{\log(1-q)} - \frac{A^2}{rR[\log(1-q)]^2} + \frac{R}{r}\right) q^{ij} \\ &+ \frac{1}{(1-q)^2}\left(-\frac{1}{2\log(1-q)} - \frac{A^2}{2rR[\log(1-q)]^2} + \frac{R}{2r}\right. \\ &\left. - \frac{1}{2[\log(1-q)]^2} - \frac{A^2}{rR[\log(1-q)]^3}\right) q^i q^j \Bigg\}\end{aligned} \tag{48}$$

As before we have

$$E[A] = R\mu = -R\log(1-q)$$

and

$$\begin{aligned}E[A^2] &= R^2\mu^2 + rR\mu \\ &= R^2[\log(1-q)]^2 - rR\log(1-q).\end{aligned}$$

Using these values for $E[A]$ and $E[A^2]$, on taking expected values of equation (48) we find

$$H_{ij}(\boldsymbol{\alpha}) = \sum \frac{1}{(1-q)^2}\left(\frac{1}{2[\log(1-q)]^2} - \frac{R}{r\log(1-q)}\right) q^i q^j \tag{49}$$

(iii) $L^*(\boldsymbol{\alpha}) = L_3(\boldsymbol{\alpha})$ further approximation and allowance for duplicates) (See equation (6.5.5).)

Here equations (34) and (35) give

$$\frac{\partial s}{\partial \mu} = -\frac{1}{2}\left(-\frac{A^2}{rR\mu^2} + \frac{R}{r}\right)$$

$$= -\frac{1}{2}\left(-\frac{A^2}{rR[\log(1-q)]^2} + \frac{R}{r}\right)$$

and

$$\frac{\partial^2 s}{\partial \mu^2} = -\frac{A^2}{rR\mu^3}$$

$$= \frac{A^2}{rR[\log(1-q)]^3}$$

Substituting these partial derivatives in equation (45) we obtain

$$H^*_{ij}(\alpha) = \sum \left\{\frac{1}{2(1-q)}\left(-\frac{A^2}{rR[\log(1-q)]^2} + \frac{R}{r}\right) q^{ij} + \frac{1}{(1-q)^2}\left(-\frac{A^2}{2rR[\log(1-q)]^2} + \frac{R}{2r} - \frac{A^2}{rR[\log(1-q)]^3}\right) q^i q^j\right\} \tag{50}$$

Using the above values of $E[A]$ and $E[A^2]$, on taking expected values of this last equation we obtain

$$H_{ij}(\alpha) = \sum \left\{\frac{1}{2(1-q)\log(1-q)} q^{ij} + \frac{1}{(1-q)^2}\left(\frac{1}{2\log(1-q)} + \frac{1}{[\log(1-q)]^2} - \frac{R}{r\log(1-q)}\right) q^i q^j\right\} \tag{51}$$

APPENDIX 2

Equality of the moments of 'actual' and 'expected' deaths and of matrices $H(\boldsymbol{\alpha})$ and $H^*(\boldsymbol{\alpha})$ under certain conditions

A. *Maximum likelihood graduation of q-type rates using initial exposures (See § 6.2)*

Suppose that the graduation formula is

$$q_{x+b}(\boldsymbol{\alpha}) = \frac{\exp\left\{\sum_{i=1}^{s} \alpha_i p_i(x)\right\}}{1 + \exp\left\{\sum_{i=1}^{s} \alpha_i p_i(x)\right\}} \tag{1}$$

where $p_1(x), \ldots, p_s(x)$ are given independent functions of x (which do *not* depend on α), not necessarily polynomials. Note that the formula

$$q_{x+b}(\boldsymbol{\alpha}) = \mathrm{LGM}_{\boldsymbol{\alpha}}^{0,s}(x) \tag{2}$$

is a particular case of formula (1).

As in Appendix 1, for notional simplification we write q, q^i, q^{ij}, A and R for $q_{x+b}(\boldsymbol{\alpha})$, $(\partial q/\partial\alpha_i)$, $(\partial^2 q/\partial\alpha_i\partial\alpha_j)$, A_x, and R_x respectively. In addition we write E for $\exp\{\Sigma_{i=1}^{s} \alpha_i p_i(x)\}$ and p_i for $p_i(x)$. Equation (1) may then be written in the form

$$q = \frac{E}{1 + E} \tag{3}$$

$$= 1 - \frac{1}{1 + E} \tag{4}$$

from which it follows that

$$q^i = \frac{1}{(1 + E)^2} \cdot E \cdot p_i = p_i \frac{E}{1 + E} \cdot \frac{1}{1 + E}$$

Thus

$$q^i = p_i q(1 - q) \tag{5}$$

(from equations (3) and (4)).

Partial differentiation of this last equation with respect to α_j gives

$$q^{ij} = p_i(1 - 2q)q^j = p_i p_j(1 - 2q)q(1 - q) \tag{6}$$

(from equation (5)).

Consider now the likelihood function when there are no duplicates. From equation (6.2.4) this is

$$L = L_1(\boldsymbol{\alpha}) = \sum_{x=n}^{m} \{A \log q + (R - A) \log (1 - q)\} \tag{7}$$

Hence

$$\frac{\partial L}{\partial \alpha_i} = \sum_{x=n}^{m} \left\{ \frac{A}{q} - \frac{R-A}{1-q} \right\} q^i = \sum_{x=n}^{m} \frac{A - Rq}{q(1-q)} q^i = \sum_{x=n}^{m} p_i(A - Rq) \quad (8)$$

by (5) above.

If the graduation is by maximum likelihood, then at the solution $\hat{\boldsymbol{\alpha}}$ each of the partial derivatives $(\partial L/\partial \alpha_i)$ $(i = 1, \ldots, s)$ will be zero. Our last equation therefore implies that a maximum likelihood graduation will have

$$\sum_{x=n}^{m} p_i(x) A_x = \sum_{x=n}^{m} p_i(x) R_x q_{x+b} \qquad (i = 1, 2, \ldots, s) \tag{9}$$

Suppose now that an LGM(0, s) formula is used—i.e. equation (2) applies. Then, for $i = 1, 2, \ldots, s$, $p_i(x)$ is a polynomial in x of degree $i - 1$. In particular, putting $i = 1$ in equation (9), we obtain (since $p_1(x) = c \neq 0$)

$$\sum_{x=n}^{m} A_x = \sum_{x=n}^{m} R_x q_{x+b} \tag{10}$$

Thus the total 'actual' and 'expected' deaths are equal.

Putting $i = 2$ in equation (9), we have

$$\sum_{x=n}^{m} (d + \mathrm{e}x) A_x = \sum_{x=n}^{m} (d + \mathrm{e}x) R_x q_{x+b}$$

where $\mathrm{e} \neq 0$. This last equation, in combination with equation (10), implies that

$$\sum_{x=n}^{m} x A_x = \sum_{x=n}^{m} x R_x q_{x+b}$$

Using a similar argument successively for $i = 3, \ldots, s$ we see that with a graduation formula of type LGM(0, s)

$$\sum_{x=n}^{m} x^i A_x = \sum_{x=n}^{m} x^i R_x q_{x+b} \qquad (i = 0, \ldots, s-1) \tag{11}$$

This establishes the important fact that for an LGM(0, s) graduation the first s moments of the 'actual' and 'expected' deaths are equal.

Consider now the matrices $H(\boldsymbol{\alpha})$ and $H^*(\boldsymbol{\alpha})$.

Equation (7) of Appendix 1 is

$$H^*_{ij}(\boldsymbol{\alpha}) = \sum \left\{ \left(\frac{R-A}{1-q} - \frac{A}{q} \right) q^{ij} + \left(\frac{A}{q^2} + \frac{R-A}{(1-q)^2} \right) q^i q^j \right\}$$
$$= \sum \left\{ \frac{Rq - A}{q(1-q)} q^{ij} + \frac{Rq^2 + A(1-2q)}{q^2(1-q)^2} q^i q^j \right\}$$

Substituting the values of q^{ij}, q^i, and q^j (from equations (5) and (6)) in this last equation, we obtain

$$H_{ij}^*(\boldsymbol{\alpha}) = \Sigma\{(Rq - A)p_i p_j(1 - 2q) + [Rq^2 + A(1 - 2q)]p_i p_j\}$$
$$= \Sigma p_i p_j Rq(1 - q)$$

Thus

$$H_{ij}^*(\boldsymbol{\alpha}) = \sum_{x=n}^{m} p_i(x)p_j(x)R_x q_{x+b}(\boldsymbol{\alpha})[1 - q_{x+b}(\boldsymbol{\alpha})] \quad (12)$$

It should be noted that this last expression does *not* depend on the actual deaths $\{A_x\}$.

Equation (8) of Appendix 1 is

$$H_{ij}(\boldsymbol{\alpha}) = \sum \frac{R}{q(1 - q)} q^i q^j = \sum Rp_i p_j q(1 - q)$$

by equation (5) above.

Hence, with a graduation formula given by equation (1) (and in particular for an LGM(0, *s*) formula) it follows that

$$H_{ij}^*(\boldsymbol{\alpha}) = H_{ij}(\boldsymbol{\alpha})$$

i.e. the matrices $H^*(\boldsymbol{\alpha})$ and $H(\boldsymbol{\alpha})$ are equal.

B *Maximum likelihood graduation of* μ*-type or m-type rates using central exposures. (See §§ 6.3 and 6.4)*

(For brevity we consider only a graduation of μ-type rates, but our remarks can readily be extended to cover the case of *m*-type rates.)

Suppose that the graduation formula is

$$\mu_{x+b}(\boldsymbol{\alpha}) = \exp\left\{\sum_{i=1}^{s} \alpha_i p_i(x)\right\} \quad (14)$$

where, as before, $p_1(x), \ldots, p_s(x)$ are given independent functions of x (which do *not* depend on $\boldsymbol{\alpha}$), not necessarily polynomials. In particular, if

$$\mu_{x+b}(\boldsymbol{\alpha}) = \mathrm{GM}_{\boldsymbol{\alpha}}^{(0,s)}(x) \quad (15)$$

equation (14) is valid.

Consider now the likelihood function when there are no duplicates. From equation (6.3.5) this is

$$L = L_1(\boldsymbol{\alpha}) = \sum_{x=n}^{m} (-R\mu + A\log\mu) \quad (16)$$

where, as before, we write μ for $\mu_{x+b+1/2}(\boldsymbol{\alpha})$.

Writing equation (14) in the form

$$\mu = E$$

we obtain, on differentiating with respect of α_i,

$$\mu^i = p_i E = p_i \mu \tag{17}$$

and, by further differentiation,

$$\mu^{ij} = p_i \mu^j = p_i p_j \mu \tag{18}$$

Equation (16) implies that

$$\frac{\partial L}{\partial \alpha_i} = \sum_{x=n}^{m} \left(- R + \frac{A}{\mu} \right) \mu^i = \sum_{x=n}^{m} p_i(- R\mu + A) \tag{19}$$

(on substitution for μ^i from equation (17)).

At the maximum likelihood solution $\hat{\boldsymbol{\alpha}}$, $\partial L/\partial \alpha_i = 0$ $(i = 1, \ldots, s)$,

Hence in a maximum likelihood graduation with the graduation formula given by equation (14) above, we have

$$\sum_{x=n}^{m} p_i(x)A_x = \sum_{x=n}^{m} p_i(x)R_x\mu_{x+b+1/2} \qquad (i = 1, 2, \ldots, s) \tag{20}$$

In particular, if the graduation formula is of type GM(0, s), $p_i(x)$ is a polynomial in x of degree $(i - 1)$. In this case, by using an argument almost identical to that § A above, we see that

$$\sum_{x=n}^{m} x^i A_x = \sum_{x=n}^{m} x^i R_x\mu_{x+b+1/2} \qquad (i = 0, \ldots, s - 1) \tag{21}$$

Thus, with a GM(0, s) formula for μ (or m) the first s moments of the 'actual' and 'expected' deaths are equal.

Consider now the matrices $H(\boldsymbol{\alpha})$ and $H^*(\boldsymbol{\alpha})$. Equation (26) of Appendix 1 is

$$H^*_{ij}(\boldsymbol{\alpha}) = \sum \left\{ \left(R - \frac{A}{\mu} \right) \mu^{ij} + \frac{A}{\mu^2} \mu^i \mu^j \right\}$$

Substituting the values of μ^{ij}, μ^i, and m^j (from equation (17) and (20), we obtain

$$H^*_{ij}(\boldsymbol{\alpha}) = \sum \left\{ \left(R - \frac{A}{\mu} \right) p_i p_j \mu + \frac{A}{\mu^2} p_i \mu p_j \mu \right\} = \sum p_i p_j R\mu$$

Thus

$$H^*_{ij}(\boldsymbol{\alpha}) = \sum_{x=n}^{m} p_i(x) p_j(x) R_x \mu_{x+b+1/2}(\boldsymbol{\alpha}) \tag{22}$$

Note that this does *not* depend on the actual deaths $\{A_x\}$.
Equation (27) of Appendix 1 is

$$H_{ij}(\boldsymbol{\alpha}) = \sum \frac{R}{\mu} \mu^i \mu^j = \sum \frac{R}{\mu} p_i \mu p_j \mu = \sum p_i p_j R\mu$$

Thus

$$H_{ij}(\boldsymbol{\alpha}) = H^*_{ij}(\boldsymbol{\alpha})$$

Hence, if the graduation formula is of the type given by equation (14), the matrices $H(\boldsymbol{\alpha})$ and $H^*(\boldsymbol{\alpha})$ are equal. In particular, for a formula of type GM(0, s) these matrices are equal.