B.A.J. 5, I, 197-236 (1999)

NON-PARAMETRIC LIKELIHOOD ENHANCEMENTS TO PARAMETRIC GRADUATIONS

BY R. J. THOMSON

ABSTRACT

Parametric graduation may fail to achieve satisfactory results without overparameterisation. Whittaker-Henderson graduation tends to constrain the graduated values towards a low-order polynomial. Non-parametric methods do not generally make direct use of true likelihood functions. This paper suggests a method of enhancing the likelihood of a parametric graduation by means of non-parametric methods, thus reducing the disadvantages of both methods.

The parametric graduation is taken to be ideally smooth by definition and is adjusted by using constrained maximum likelihood estimation to obtain better fidelity to the experience. The constraint imposes a minimum sacrifice of smoothness, in terms of a quantitative smoothness criterion, from the initial ideal. The method is not entirely objective in that, in some cases, professional judgement is required in order to assess the degree of smoothness that can be imposed. In other cases the method provides an objective optimum. In either case, by quantifying the degrees of departure from perfect fidelity and from ideal smoothness, the suggested method provides useful and theoretically sound criteria for the purposes of the optimisation process. In particular, by inverting the parametric graduation formula for the purposes of defining the smoothness criterion, the method ensures that the smoothness criterion is consistent over the whole age range, thus resolving the main objection to non-parametric graduation.

The method is applied to the 1979-82 experience for life office pensioners in the United Kingdom with positive results.

KEYWORDS

Parametric Graduation; Whittaker-Henderson Graduation; Non-Parametric Graduation

CONTACT ADDRESS

R. J. Thomson, B.Sc., F.I.A., Department of Statistics & Actuarial Science, University of the Witwatersrand, Private Bag 3, WITS 2050, South Africa, Tel: +27(0)11-646-5332; Fax: +27(0)11-339-6640; E-mail: rthomson@global.co.za

1. INTRODUCTION

1.1 Suppose that a mortality investigation has been made. Let R, with ith component R_i, denote an n-component vector of life-years of exposure to risk during the age intervals $[1,2),\ldots,[i,i+1),\ldots,[n,n+1)$, where, for convenience, the ages are numbered from 1 to n. Let A, with ith component A_i, denote the corresponding numbers of deaths observed. The components of R must be central or initial exposed to risk, according as annual average forces or annual rates of mortality are to be graduated. If the components of R are initial exposed to risk and annual average forces of mortality are to be graduated, then for each i we substitute R_i with:

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

$$R_i - \frac{A_i}{2}. \tag{1}$$

If they are central exposed to risk and annual rates of mortality are to be graduated, then for each i we substitute R_i with:

$$R_i + \frac{A_i}{2}. \tag{2}$$

1.2 In this paper, consideration is given to the graduation both of annual average forces of mortality:

$$\bar{\mu}_x = \int_0^1 \mu_{x+t}\, dt \tag{3}$$

where μ_x is the force of mortality at age x, and of annual rates of mortality:

$$q_x = 1 - \exp(-\bar{\mu}_x). \tag{4}$$

The quantities $\bar{\mu}_x$ and q_x are referred to as 'probability measures'. They represent the true underlying measures, as opposed to the graduated estimates to be obtained. The reasons for the estimation of $\bar{\mu}_x$ instead of $\mu_{x+1/2}$ are explained in ¶2.3.

1.3 Let y, with ith component y_i, denote an n-component vector of estimates of the probability measures to be graduated.

1.4 Let F: $\mathbb{R}^n \to \mathbb{R}$ denote a function:

$$F(y) = -\kappa \log L(y) - k \tag{5}$$

where L is the likelihood function for the probability measure concerned and $\kappa > 0$ and k are constants (independent of y) such that, for y^* equal to the maximum-likelihood estimate of y:

$$F(y^*) = 0. \tag{6}$$

F is thus a non-negative function of y, which attains a minimum value of zero when y is equal to the maximum-likelihood estimate of the probability measure concerned. In order to maximise the likelihood of y, it is therefore necessary to minimise $F(y)$. The definition of $F(\cdot)$ is dealt with in Section 2. The reason for using $F(y)$ instead of $\log L(y)$ is that it avoids unnecessary coefficients in the latter and unnecessary leading digits in its values. In practice, the author has found this helpful in numerical applications.

1.5 Let the n-component vector $\tilde{v}$, with xth component:

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

$$\tilde{v}_x = \frac{A_x}{R_x} \tag{7}$$

denote the crude estimate of the probability measure, and let the n-component vector u, with xth component:

$$u_x = \phi(x \mid \alpha)$$

denote a parametric graduation, where α is a vector of parameters. Let $S(y)$ denote a non-negative measure of the smoothness of y such that, for a criterion defined with reference to $\phi(\cdot|\alpha)$:

(1) $S(u)=0$; and

(2) $S(y^{**}) > S(y^{*})$ means that y^{*} is smoother than y^{**} on that criterion.

In effect, therefore, the parametric graduation u is perfectly smooth on the criterion of the function used in its derivation, and $S(y)$ denotes the extent to which y departs from the smoothness of u. The definition of $S(\cdot)$ is dealt with in Section 3.

1.6 The graduation problem may be generalised in terms of the classical constrained optimisation problem (e.g. Walsh, 1975) as follows:

$$\text{minimise } F(y) \text{ subject to } S(y)=s \tag{8}$$

where:

$$0 \leq s \leq S(\tilde{v}).$$

The graduation problem is thus expressed as a maximum-likelihood estimation problem, subject to the constraint of a criterion of smoothness. The problem is solved in Section 4.

1.7 In Section 5 the method is applied to a specific set of parametric graduations. It is shown that the method achieves marked improvements in certain graduations. Section 6 concludes.

1.8 A similar specification was considered by Chan, Chan & Yu (1984), except that:

(1) they defined $F(\cdot)$ as:

$$F(y) = \sum_{x=1}^{n} w_x \mid \tilde{v}_x - y_x \mid^b$$

which, with appropriately determined values of w_x and with $b=2$, may approximate a monotonically decreasing function of the likelihood of y; and

(2) they defined $S(\cdot)$ as:

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

$$S(y) = \sum_{x=1}^{n-z} | \Delta^z y_x |^b \tag{9}$$

which would conform to the specification of ¶1.3 only if $\phi(x|\alpha)$ is a polynomial of order z.

It was noted that, for $b=2$, this specification reduced to the Whittaker-Henderson Type B graduation method.

1.9 For convenience, reference is made in this paper to 'mortality'. Although some of the parametric formulae mentioned may be more applicable to mortality, the principles are equally applicable to other contingencies.

1.10 A SAS program has been written to apply the methodology of this paper, and is available from the author. Some of the formulae in the paper are more for the benefit of users of that program, which is cross-referenced to the paper, than of other readers; the author requests the indulgence of the reader.

2. THE LIKELIHOOD FUNCTION

2.1 The Whittaker-Henderson Type B method uses the goodness-of-fit function:

$$\sum_{x=1}^{n} w_x(\tilde{v}_x - y_x)^2. \tag{10}$$

In the application of this method it is common (London, 1985, p54), when graduating annual rates of mortality, to define:

$$w_x = \frac{R_x}{\tilde{v}_x(1-\tilde{v}_x)} \tag{11}$$

that is to use goodness-of-fit weightings equal to the inverse of the initial estimates of the variance. This may be justified on the grounds that, for large values of R_x, the observed rates of decrement are approximately normally distributed, so that minimising the value of (10) with w_x defined as in equation (11) is approximately equivalent to maximising the likelihood.

2.2 However, there are two approximations involved here. First, as noted by London, it would be preferable to use:

$$w_x = \frac{R_x}{v_x(1-v_x)} \tag{12}$$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

where v_x is the xth component of the vector of graduated values; that is, to use goodness-of-fit weightings equal to the inverse of the final estimates of the variance. Secondly, the likelihood function implied by equation (12) is only asymptotically correct.

2.3 In fact, if we are graduating average annual forces of mortality then, for a given value of v_x — say y_x — the number of deaths at age x follows a Poisson distribution. As noted by Forfar, McCutcheon & Wilkie (1988, pp29,30), the log-likelihood function is:

$$\log L(y) = \sum_{x=1}^{n} \{-R_x y_x + A_x \log(R_x y_x) - \log(A_x!)\} \tag{13}$$

Those authors approximate the integral in equation (3) by the value $\mu_{x+1/2}$. However, that approximation is unnecessary. In fact, if q_x is to be found, equation (4) may be used to obtain it direct from $\bar{\mu}_x$, so that the approximation would only serve to complicate the calculations.

2.4 If, in equation (5), we let:

$$\kappa = 1 \tag{14}$$

$$k = \sum_{x=1}^{n} k_x$$

and

$$k_x = \begin{cases} \log(A_x!) + A_x(1 - \log A_x) & \text{for } A_x > 0 \\ 0 & \text{for } A_x = 0 \end{cases} \tag{15}$$

then it may be shown that:

$$F(y) = \sum_{x=1}^{n} (R_x y_x - h_x) \tag{16}$$

where:

$$h_x = \begin{cases} A_x\left(1 + \log \dfrac{y_x}{\tilde{v}_x}\right) & \text{for } \tilde{v}_x > 0 \\ 0 & \text{otherwise} \end{cases}$$

satisfies the requirements of ¶1.4. To prove this, we note, by substituting equations (13) to (15) in (5), that:

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

$$F(y) = \sum_{x=1}^{n} f_x$$

where:

$$f_x = R_x y_x - A_x \log(R_x y_x) + \log(A_x!) - \log(A_x!) - A_x(1 - \log A_x)$$

$$= \begin{cases} R_x y_x - A_x\left(1 + \log \dfrac{R_x y_x}{A_x}\right) & \text{for } A_x > 0 \\ R_x y_x & \text{otherwise.} \end{cases}$$

Thus, from equation (7):

$$f_x \begin{cases} R_x y_x - A_x\left(1 + \log \dfrac{y_x}{\tilde{v}_x}\right) & \text{for } \tilde{v}_x > 0 \\ R_x y_x & \text{otherwise} \end{cases}$$

from which equation (16) follows. Now $\tilde{v}$ is the maximum-likelihood estimate of the probability measure (Forfar, McCutcheon & Wilkie, 1988, p7). Finally, for $y_x = \tilde{v}_x$:

$$f_x = \begin{cases} R_x \tilde{v}_x - A_x = 0 & \text{for } \tilde{v}_x > 0 \\ R_x \tilde{v}_x = 0 & \text{otherwise} \end{cases}$$

so that equation (6) holds for $y^* = \tilde{v}$.

2.5 Similarly, if we are graduating annual rates of mortality, then the number of deaths at age x is approximately binomially distributed and $F(\cdot)$ may be defined as:

$$F(y) = \sum_{x=1}^{n} \left\{ (R_x - A_x) \log\left(\frac{1 - \tilde{v}_x}{1 - y_x}\right) - h_x \right\} \tag{17}$$

where:

$$h_x = \begin{cases} A_x \log\left(\dfrac{y_x}{\tilde{v}_x}\right) & \text{for } \tilde{v}_x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

The approximation arises from the fact that not all lives are observed from exact age x to exact age $x+1$ or earlier death.

2.6 Equations (16) and (17) also presuppose that there are no duplicates. As shown by Forfar, McCutcheon & Wilkie (1988, pp28,32), the presence of duplicates may be allowed for by using normal approximations. Let:

$$\rho_x = \frac{\sum_i i^2 f_x^i}{\sum_i i f_x^i}$$

where $f_x^i (i = 1, 2, \ldots)$ denotes the proportion of lives that are included i times, so that:

$$\sum_i f_x^i = 1.$$

If we are graduating average annual forces of mortality then:

$$\log L(y) = -\frac{1}{2}\sum_{x=1}^{n}\left\{\log(2\pi\rho_x R_x y_x) + \frac{(A_x - R_x y_x)^2}{\rho_x R_x y_x}\right\}.$$

Thus:

$$\frac{\partial}{\partial y_x}\log L(y) = -\frac{1}{2}\left(\frac{1}{y_x} + \frac{R_x^2 y_x^2 - A_x^2}{\rho_x R_x y_x^2}\right). \tag{18}$$

By setting the right-hand side of equation (18) equal to zero, we find that the maximum-likelihood estimator of y_x is:

$$y_x^* = \frac{\sqrt{A_x^2 + \frac{1}{4}\rho_x^2} - \frac{1}{2}\rho_x}{R_x}. \tag{19}$$

Now it is clear from equation (19) that y_x^* is a biased estimate of y_x; in fact it understates y_x. On the other hand, as R_x increases with constant $\tilde{v}_x$ (so that, from equation (7), A_x increases proportionately to R_x), it is evident that y_x^* is asymptotically unbiased. Nevertheless, for finite values of R_x, it is desirable to eliminate the bias.

2.7 For this purpose, following Forfar, McCutcheon & Wilkie (1988, pp32, 33), we use the normalised variable:

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

$$z_x = \frac{A_x - R_x y_x}{\sqrt{\rho_x R_x y_x}}$$

so that the likelihood function is such that:

$$\log L(z) = -\frac{1}{2}\sum_{x=1}^{n} z_x^2.$$

The maximum-likelihood estimator of z_x is:

$$z_x^* = 0$$

so that:

$$z_x = z_x^* \Rightarrow y_x = \frac{A_x}{R_x} = \tilde{v}_x$$

which is unbiased. Thus, setting $\kappa = 2$ and $k = 0$, we may define:

$$F(y) = \sum_{x=1}^{n} \frac{(A_x - R_x y_x)^2}{\rho_x R_x y_x}$$

that is, the χ^2 test statistic, noting that:

$$F(\tilde{v}) = 0.$$

2.8 The use of the normal approximation in ¶¶2.6 and 2.7 is reasonable only if the expected number of deaths at each age is sufficiently large. Forfar, McCutcheon & Wilkie (1988, p42) suggest a minimum of 5. In order to overcome this problem, they propose a method of grouping of neighbouring ages, which is used in this paper. This grouping method may be achieved by means of an $n^* \times n$ matrix P such that:

$$E^* = PE$$

gives a grouped vector of the expected deaths on the basis of the original graduation. The same matrix may then be multiplied by the vector of actual deaths to give:

$$A^* = PA \tag{20}$$

the grouped vector of actual deaths. It may also be multiplied by the vector of expected deaths on the regraduation to give:

$$E^*(y) = PE(y) \tag{21}$$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

and by the vector of variances σ^2 with typical component $\sigma_x^2 = \rho_x R_x y_x$ to give $\sigma^{2\clubsuit} = P\sigma^2$. The age range of each group may be determined by multiplying the age vector by a matrix derived from P.

2.9 If we are graduating annual average forces, then the definition of $F(\cdot)$ becomes:

$$F(y) = \sum_{j=1}^{n^\clubsuit} \frac{\left\{\sum_{i=1}^{n} p_{ji}(A_i - R_i y_i)\right\}^2}{\sum_{i=1}^{n} p_{ji}\rho_i R_i y_i} \tag{22}$$

$$= \sum_{j=1}^{n^\clubsuit} z_j^2 \tag{23}$$

where:

$$z_j = \frac{A_j^\clubsuit - E_j^\clubsuit(y)}{\sigma_j^\clubsuit} \tag{24}$$

$$\sigma^{2\clubsuit} = P\sigma^2 \tag{25}$$

and

$$\sigma_x^2 = \rho_x R_x y_x. \tag{26}$$

2.10 Similarly, if we are graduating annual rates of mortality, then:

$$F(y) = \sum_{j=1}^{n^\clubsuit} \frac{\left\{\sum_{i=1}^{n} p_{ji}(A_i - R_i y_i)\right\}^2}{\sum_{i=1}^{n} p_{ji}\rho_i R_i y_i (1 - y_i)} \tag{27}$$

$$= \sum_{j=1}^{n^\clubsuit} z_j^2$$

where:

$$z_j = \frac{A_j^\clubsuit - E_j^\clubsuit(y)}{\sigma_j^\clubsuit}$$

$$\sigma^{2\clubsuit} = P\sigma^2$$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

and

$$\sigma_x^2 = \rho_x R_x y_x(1-y_x). \tag{28}$$

3. SMOOTHNESS

3.1 As noted by London (1985, pp6, 62) the use of equation (9) constrains v_x towards a polynomial of order z. This is not necessarily appropriate. Other typical functions have been identified for probability measures. For the purposes of mortality investigations, Forfar, McCutcheon & Wilkie (1988, pp15-17) suggest a 'Gompertz-Makeham formula of type (r, s)':

$$GM_\alpha^{r,s}(x) = \sum_{i=1}^{r} \alpha_i x^{i-1} + \exp\left\{ \sum_{i=r+1}^{r+s} \alpha_i x^{i-r-1} \right\} \tag{29}$$

or a 'Logit Gompertz-Makeham formula of type (r, s)':

$$LGM_\alpha^{r,s}(x) = \frac{GM_\alpha^{r,s}(x)}{1 + GM_\alpha^{r,s}(x)}. \tag{30}$$

In particular, they suggest the former for the graduation of forces of mortality and the latter for rates of mortality.

3.2 In practice, numerous forms of functions have been used for parametric graduations. In general, we may express any such function as a function of polynomials in x, so that:

$$\phi(x \mid \alpha) \equiv \Phi(t_1, \ldots, t_m \mid \beta)$$

where t_γ is an n-component vector with xth component:

$$t_{\gamma,x} = \sum_{j=1}^{v(\gamma)} \beta_{\gamma,j} x^{j-1} \quad \text{for } \gamma = 1, \ldots, m \tag{31}$$

$v(\gamma)$ is the order of the γth polynomial and m is the smallest number of polynomials in terms of which $\phi(x \mid \alpha)$ can be expressed. At worst, $\beta_{\gamma,j} \equiv \alpha_\gamma$, in which case every polynomial comprises only one term; otherwise the transformation will help to simplify the process as explained below. For example, for a GM formula, we have, from equation (29):

$$\Phi_x(t_1, t_2) = t_{1,x} + e^{t_{2,x}} \tag{32}$$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

$$\beta_{1,j} = \alpha_j \quad \text{for} \quad j = 1,\ldots, r \tag{33}$$

$$\beta_{2,j-r} = \alpha_j \quad \text{for} \quad j = r+1,\ldots, r+s. \tag{34}$$

For an LGM formula, we have, from equation (30):

$$\Phi_x(t_1, t_2) = \frac{t_{1,x} + e^{t_{2,x}}}{1 + t_{1,x} + e^{t_{2,x}}} \tag{35}$$

with $\beta_{\gamma,j}$ as before.

3.3 In order to constrain y_x towards u_x, we define our smoothness function as:

$$S(y) = \sum_{\gamma=1}^{m} \eta_\gamma S_\gamma(y) \tag{36}$$

where:

$$S_\gamma(y) = \sum_{x=1}^{n-z(\gamma)} |\Delta^{z(\gamma)} g_\gamma(y_x)|^b \quad \text{and} \quad b > 1 \tag{37}$$

and the value of the parameter b (referred to below as the 'power' of the smoothness function), and for each γ the values of the parameters η_γ (referred to below as the 'weightings of the components' of the smoothness function) and $z(\gamma)$ (referred to below as the 'orders of differencing' of the smoothness function), need to be specified by the user. The rationale underlying the formulation of equation (37) is explained below. In that equation, $g_\gamma(y_x)$ is the inverse of Φ with respect to t_γ, so that:

$$g_\gamma(\Phi(t_1,\ldots,t_m)) = t_\gamma \quad \text{for} \quad \gamma = 1,\ldots, m.$$

For example, for a GM formula we have:

$$g_1(y_x) = y_x - e^{t_2} \quad \text{and} \quad g_2(y_x) = \log(y_x - t_1) \tag{38}$$

and, for an LGM formula, we have:

$$y_x = \frac{t_1 + e^{t_2}}{1 + t_1 + e^{t_2}}$$

so that:

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

$$t_1 + e^{t_2} = \frac{y_x}{1 - y_x}$$

and hence:

$$g_1(y_x) = \frac{y_x}{1 - y_x} - e^{t_2} \quad \text{and} \quad g_2(y_x) = \log\left(\frac{y_x}{1 - y_x} - t_1\right). \tag{39}$$

3.4 Since $t_{\gamma,x}$ is a polynomial of order $\nu(\gamma)$ (i.e. of degree $\nu(\gamma)$–1), its $\nu(\gamma)$th differences are zero. Thus, if $z(\gamma) \geq \nu(\gamma)$ for $\gamma = 1,\ldots,m$, then $S_\gamma(u) = 0$ for $\gamma = 1,\ldots, m$, so that $S(u) = 0$. Furthermore, provided that, for $\gamma = 1,\ldots,m$, $\eta_\gamma \geq 0$, it may be seen from equations (36) and (37) that $S(y) \geq 0$. The use of $S(\cdot)$ as a smoothing function is therefore consistent with the requirements of ¶1.5, provided $z(\gamma) \geq \nu(\gamma)$ for $\gamma = 1,\ldots, m$. If the order of one of the polynomials is particularly high (say greater than 4), it may be thought that this implies too high an order of differencing. However, if the parametric graduation is sufficiently smooth, the high order of differencing will be both feasible (in the sense that it will not be imposing an impossibly high-order polynomial) and desirable (in the sense that it will be very sensitive to any roughness). Otherwise the parametric graduation should be reconsidered.

3.5 In particular, for a GM or LGM formula, we have $\nu(1) = r$ and $\nu(2) = s$. It is therefore necessary that $z(1) \geq r$ and $z(2) \geq s$, where, in general, it is appropriate to apply the equalities. In this way, if a polynomial of order r is considered smooth, the minimisation of $S(y)$ would constrain the probability measure towards smoothness in the desired sense.

3.6 From the above, it may be seen that the purpose of inverting Φ is to convert the partly graduated values y_x into functions that will gradually be constrained towards polynomials of order $z(\gamma)$. It is thus in the ranges of the polynomials that the criterion of smoothness is applied. It is assumed that (as is generally the case with GM and LGM formulae), if the polynomials $t_{\gamma,x}$ are smooth, the probability measure $\Phi(t_1,\ldots, t_m|\beta)$ will also be smooth. The advantage of applying the smoothness criterion in the ranges of the polynomials, instead of the range of the probability measure, is that the scale of the smoothness in the ranges of the polynomials is unaffected by the transcendental functions that Φ applies to them. In the case of mortality, the transcendental function is generally, as in the GM formula, an exponential function, whose purpose is to facilitate graduation, both at young ages and at old ages, where the scales of the range of the probability measure are entirely different. It therefore seems preferable to apply the smoothness criterion before exponentiation.

3.7 The parameter b provides control over the emphasis to be placed on large absolute departures from smoothness relative to small ones. As b increases, relatively less emphasis is placed on smaller departures. In fact, as b tends to infinity, the criterion tends to focus on the largest departure, thus reducing the

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

problem to a minimax problem (i.e. to the minimisation of the maximum departure). For the applications made in this paper, the author found that values of 2 or 4 gave satisfactory results. For even values of b the use of absolute values is unnecessary, but intermediate (including non-integral) values may be useful, and the formulation has been generalised for that purpose.

3.8 The values assigned to $\{\eta_\gamma\}$ will depend on the nature of the graduation formula. If, for example, t_1 is designed to cater for an accident hump over a restricted range of ages and t_2 for senescent mortality over the rest of the range, then we could set $\eta_1 = 0$ and $\eta_2 = 1$. The specification of the smoothness function thus enables the user to control the relative emphases given to the polynomials in terms of which the parametric graduation is specified. If there is no apparent restriction of the effect of any t_γ relative to the others, then each could be assigned a value. Appropriate values may be:

$$\eta_1 = \frac{|\alpha_{r+s}|}{|\alpha_r| + |\alpha_{r+s}|} \quad \text{and} \quad \eta_2 = \frac{|\alpha_r|}{|\alpha_r| + |\alpha_{r+s}|}.$$

With these values, the weightings applied to the highest-powered term of each polynomial will be inversely proportionate to the absolute value of its coefficient, and the sum of the weightings will be 1.

4. SOLUTION

4.1 For a function f: $\mathbb{R}^n \to \mathbb{R}$, let D denote the partial differentiation operator, so that $Df(y)$ is an n-component vector with typical component $\frac{\partial}{\partial y_i} f(y)$.

4.2 The solution to the problem (8) (Walsh, 1975 pp6,7) satisfies the equations:

$$DF(v) + \lambda DS(v) = 0$$

and

$$S(v) = s.$$

It may be noted that this is similar in form to the Whittaker-Henderson specification, except that s is predetermined instead of λ. (Also, $F(\cdot)$ and $S(\cdot)$ are differently defined.)

4.3 If we define:

$$M(y, \lambda) = F(y) + \lambda S(y) \tag{40}$$

then the problem reduces to:

$$DM(v,\lambda) = DF(v) + \lambda DS(v) = 0 \tag{41}$$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

and

$$S(v) = s$$

provided $D(DM(v))^{\mathsf{T}}$ is a positive-definite matrix.

4.4 The ith component of $DF(y)$ may be obtained from equations (16), (17), (22) and (27) as follows:

(1) in the absence of duplicates:

— for average annual forces of mortality:

$$R_i - \frac{A_i}{y_i} \tag{42}$$

— for annual rates of mortality:

$$\frac{R_i - A_i}{1 - y_i} - \frac{A_i}{y_i} \tag{43}$$

(2) in the presence of duplicates:

— for average annual forces of mortality:

$$-\sum_{j=1}^{n^*} \zeta_j (2 + \rho_i \zeta_j) p_{ji} R_i \tag{44}$$

where:

$$\zeta_j = \frac{\sum_{i=1}^{n} p_{ji}(A_i - R_i y_i)}{\sum_{i=1}^{n} p_{ji} \rho_i R_i y_i} \tag{45}$$

— for annual rates of mortality:

$$-\sum_{j=1}^{n^*} \zeta_j \{2 + \rho_i \zeta_j (1 - 2y_i)\} p_{ji} R_i \tag{46}$$

where:

$$\zeta_j = \frac{\sum_{i=1}^{n} p_{ji}(A_i - R_i y_i)}{\sum_{i=1}^{n} p_{ji} \rho_i R_i y_i (1 - y_i)}. \tag{47}$$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

4.5 From ¶3.3 we have:

$$S(y)=\sum_{\gamma=1}^{m}\eta_\gamma\sum_{x=1}^{n-z(\gamma)}|\Delta^{z(\gamma)}g_\gamma(y_x)|^b \quad \text{where} \quad b>1 \quad \text{(equations (36) and (37)).}$$

Let $K(z(\gamma))$, with typical element $k_{ij}^{(\gamma)}$ denote the $(n-z(\gamma))\times n$ $z(\gamma)$th difference matrix, so that:

$$K(z(\gamma))y=\begin{vmatrix}\Delta^{z(\gamma)}y_1\\ \cdot\\ \cdot\\ \Delta^{z(\gamma)}y_{n-z(\gamma)}\end{vmatrix}.$$

For example, for $z(\gamma)=3$:

$$K(z(\gamma))=\begin{vmatrix}-1 & 3 & -3 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0\\ 0 & -1 & 3 & -3 & 1 & \cdots & 0 & 0 & 0 & 0 & 0\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdot & \cdot\\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 3 & -3 & 1\end{vmatrix}.$$

For computational purposes, this matrix may be determined by truncating the first $z(\gamma)-1$ rows of the matrix:

$$\{K(1)\}^{z(\gamma)}. \tag{48}$$

Let $c^{(\gamma)}$ be the n-component vector with xth component $c_x^{(\gamma)}=g_\gamma(y_x)$.
Then:

$$\Delta^{z(\gamma)}g_\gamma(y_x)=\sum_{i=1}^{n}k_{xi}^{(\gamma)}c_i^{(\gamma)}.$$

Let:

$$\operatorname{sgn}(y)=\begin{cases}-1 & \text{if } y<0\\ 0 & \text{if } y=0\\ 1 & \text{if } y>0.\end{cases} \tag{49}$$

Then:

$$|\Delta^{z(\gamma)}g_\gamma(y_x)|=\operatorname{sgn}\left(\sum_{i=1}^{n}k_{xi}^{(\gamma)}c_i^{(\gamma)}\right)\sum_{i=1}^{n}k_{xi}^{(\gamma)}c_i^{(\gamma)}.$$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

Let:

$$c_x^{(\gamma)\prime} = \frac{\partial}{\partial y_x} g_\gamma(y_x).$$

Then:

$$\frac{\partial}{\partial y_j} S(y) = \sum_{\gamma=1}^{m} \eta_\gamma \frac{\partial}{\partial y_j} \left\{ \sum_{x=1}^{n-z(\gamma)} | \Delta^{z(\gamma)} g_\gamma(y_x) |^b \right\}$$

$$= b \sum_{\gamma=1}^{m} \eta_\gamma \sum_{x=1}^{n-z(\gamma)} \left| \sum_{i=1}^{n} k_{xi}^{(\gamma)} c_i^{(\gamma)} \right|^{b-1} \operatorname{sgn}\left(\sum_{i=1}^{n} k_{xi}^{(\gamma)} c_i^{(\gamma)} \right) k_{xj}^{(\gamma)} c_j^{(\gamma)\prime}.$$

Let:

$$a_x^{(\gamma)} = \sum_{i=1}^{n} k_{xi}^{(\gamma)} c_i^{(\gamma)} \tag{50}$$

and

$$e_x^{(\gamma)} = \operatorname{sgn}(a_x^{(\gamma)}) \, | \, a_x^{(\gamma)} \, |^{b-1} . \tag{51}$$

Then:

$$\frac{\partial}{\partial y_j} S(y) = b \sum_{\gamma=1}^{m} \eta_\gamma \sum_{x=1}^{n-z(\gamma)} k_{xj}^{(\gamma)} e_x^{(\gamma)} c_j^{(\gamma)\prime}. \tag{52}$$

4.6 The value of $c_x^{(\gamma)\prime}$ depends on the definition of $g_\gamma(\cdot)$. For example, for a GM formula we have, from equations (38):

$$c_x^{(1)\prime} = 1 \tag{53}$$

$$c_x^{(2)\prime} = \frac{1}{y_x - t_1} \tag{54}$$

and, for an LGM formula, we have, from equations (39):

$$c_x^{(1)\prime} = \frac{1}{(1 - y_x)^2} \tag{55}$$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

$$c_x^{(2)\prime} = \frac{1}{(1-y_x)\{y_x - t_1(1-y_x)\}}. \tag{56}$$

4.7 By substituting values of $c_x^{(\eta)\prime}$ into equation (52) we obtain the jth component of $DS(y)$ and, using the value of $DF(y)$ from ¶4.4, we obtain:

$$DM(v,\lambda) = DF(v) + \lambda DS(v) \quad \text{(equation (41))}.$$

4.8 $M(y,\lambda)$ is not necessarily convex over the whole of the domain to which y may belong. For this reason, it is necessary to use an optimisation method that does not depend on the convexity of the objective function and that will converge to a solution (even though it may only be a local solution), regardless of the commencing value of the argument. The Davidon-Fletcher-Powell method (e.g. Walsh, 1975, pp110-117) satisfies these criteria. This method is a quasi-Newton method; it combines a strategy which (unlike Newton's method) will converge regardless of the commencing value, with a rapidly converging strategy (like Newton's method) once convergence becomes imminent. An algorithm for the DFP formula is set out in Appendix A.

4.9 In step A.2 of the above algorithm, a one-dimensional search is required. For this purpose, Davidon's cubic interpolation method (Walsh, 1975, pp97-101) may be used. That method uses function values at two points on the line $y_k + \beta d_k$, where y_k and d_k are constant n-component vectors and β is the independent scalar variable, together with values of the directional derivatives of the function along that line at the same two points.

4.10 Because of the particular constraints of the graduation problem, Davidon's cubic interpolation method needs to be modified.

4.11 First, it is necessary to ensure that, for all values of k and x:

$$y_{k,x} > 0. \tag{57}$$

Similarly, if we are graduating rates, it is necessary that:

$$y_{k,x} < 1. \tag{58}$$

4.12 Furthermore, if a GM or LGM formula is used, there is a discontinuity in $g_2(y_x)$ for $y_x = t_{1,x}$. This may be seen from equations (38). It is therefore necessary to ensure that, for all values of k and x:

$$y_{k,x} > t_{1,x}. \tag{59}$$

Similarly, if an LGM formula is used, there are discontinuities in $g_1(y_x)$ and $g_2(y_x)$ for $y_x = 1$. This may be seen from equations (39). For an LGM formula (regardless of whether forces or rates are being graduated) it is therefore also necessary that:

$$y_{k,x} < 1. \tag{60}$$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

Furthermore, if an LGM formula is used, there is a discontinuity in $g_2(y_x)$ for $y_x = \frac{t_{1,x}}{1+t_{1,x}}$. This may be seen from equation (39). It is then also necessary that:

$$y_{k,x} > \frac{t_{1,x}}{1+t_{1,x}}. \tag{61}$$

4.13 An algorithm for Davidon's cubic interpolation method, modified to allow for these constraints, is set out in Appendix B. In that algorithm, in order to apply the above constraints, the function $e(y)$ is introduced, which is defined as follows.

Let:

$$g = \max_x \left(\delta_x \left| \frac{\tilde{v}_x - y_x}{d_{kx}} \right| \right) \tag{62}$$

where:

$$\delta_x = \begin{cases} 1 & \text{if } d_{kx} \neq 0 \\ 0 & \text{if } d_{kx} = 0 \end{cases}$$

and $d_{kx}=0$ is the xth component of the direction vector d_k defined in ¶A.1 of Appendix A. The purpose of this variable is to avoid applying constraints where they are unnecessary; g is, in effect, greater than the greatest value of β_k that might conceivably be required in ¶A.2. We then initially define $e(y)$ as:

$$\min_x \left\{ \delta_x \left| \frac{y_x}{2d_{kx}} \right| + (1-\delta_x)g \right\} \tag{63}$$

where:

$$\delta_x = \begin{cases} 1 & \text{if } d_{kx} < 0 \\ 0 & \text{if } d_{kx} \geq 0 \end{cases}$$

so that, if $\beta_k \leq e(y)$, constraint (57) will be satisfied. The value of $e(y)$ is then made subject to a maximum of:

$$\min_x \left\{ \delta_x \left| \frac{y_x - t_{1,x}}{2d_{kx}} \right| + (1-\delta_x)g \right\} \tag{64}$$

with the same definition of δ_x, so that, if $\beta_k \leq e(y)$, constraint (59) will be

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

satisfied. If we are graduating rates, or if an LGM formula is being used, the value of $e(y)$ is then further subjected to a maximum of:

$$\min_x \left\{ \delta_x \left| \frac{1-y_x}{2d_{kx}} \right| + (1-\delta_x)g \right\} \tag{65}$$

where:

$$\delta_x = \begin{cases} 1 & \text{if } d_{kx} > 0 \\ 0 & \text{if } d_{kx} \leq 0 \end{cases}$$

so that, if $\beta_k \leq e(y)$, constraints (58) and (60) will be satisfied. Finally, if an LGM formula is being used, the value of $e(y)$ is subjected to a maximum of:

$$\min_x \left\{ \delta_x \left| \frac{t_{1,x} - (1+t_{1,x})y_x}{2d_{kx}(1+t_{1,x})} \right| + (1-\delta_x)g \right\} \tag{66}$$

with δ_x defined as for the previous maximum, so that, if $\beta_k \leq e(y)$, constraint (61) will be satisfied.

4.14 The algorithm referred to in ¶4.8 may be used to solve equation (41) for a given value of λ. Values of λ for particular values of $s = S(v)$ may then be found by successive approximation. For each value of s the minimum value of $F(v)$ may thus be found. The value of v at that minimum is the local solution to the graduation problem as specified in ¶1.6.

4.15 The χ^2 test statistic may be used as a criterion for selection of an appropriate regraduation. Let χ^2_y and $p(\chi^2_y)$ denote, respectively, the chi-squared test statistic for the estimate vector y, and the probability that the value of that statistic will exceed χ^2_y. (Note that the subscript of χ^2 denotes the vector for which the statistic is determined, not the number of degrees of freedom of that statistic.) The number of degrees of freedom of χ^2_u is:

$$n^* - l_\alpha \tag{67}$$

where l_α is the number of components in the parameter vector α. For any regraduation the number of degrees of freedom is 0, because, by virtue of equation (41), there are n linear constraints in the determination of v. In fact, it may be easily verified that, for a given regraduation, there is only one possible set of observed deaths that could have produced that regraduation, and that is the set $\{A_i\}$. Every regraduation thus fully utilises all the information contained in the data; $p(\chi^2_v)$ should, therefore, be interpreted as the probability that the value of the chi-squared test statistic would have exceeded χ^2_v if v had been a parametric graduation with the same number of parameters as u.

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

4.16 In general, the range of values of $s = S(v)$ that would be considered is that for which:

$$p(\chi_u^2) < p(\chi_v^2) \leq 0.5. \tag{68}$$

Lower values of $p(\chi_v^2)$ would not constitute an enhancement of likelihood; greater values would suggest overfitting. If the value of s for which $p(\chi_v^2)=0.5$ is satisfactory and other tests give satisfactory results, and if, on inspection, the regraduation for that value of s is acceptable, then that regraduation should be adopted. Otherwise lower values of s should be considered, until an acceptable regraduation is found.

4.17 One of the major criteria in considering a particular regraduation is its shape. For example, in the context of male normal or late retirement pensioners, Forfar, McCutcheon & Wilkie (1988, p84) define a good shape as being:

> "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 reasons for poor shape may be paucity of data at the extreme ages or distortions of the sample observed. It should be noted, however, that the imposition of criteria of shape, whether in the parametric graduation itself or only in the enhancement, may result in a suboptimal graduation merely because of concerns about the extreme ages, at which the data are relatively scanty. Under these circumstances, if, in practice, mortality rates are required at the extreme ages, it may be better to obtain them with reference to other tables than to rely on the experience. If such a procedure is to be adopted, it may be unnecessary to worry about the shape of the regraduation. Otherwise it may be necessary to impose more stringent smoothness constraints.

5. APPLICATION

5.1 *Data used*

5.1.1 The 1979-82 experience for male normal or late retirement life office pensioners in the United Kingdom on the basis of lives with central exposed to risk are shown in Forfar, McCutcheon & Wilkie (1988, p92). The corresponding experience for females normal or late retirement and males and females early retirement with initial exposed to risk was supplied to the author by the Secretary of the CMI Bureau. The results of the graduation of the force of mortality from those experiences are shown in CMI Committee (1988, pp78-81). Rates of mortality based on the forces obtained from the graduations for normal or late retirement were published in CMI Committee (1990) as standard tables, which are designated PML80Base and PFL80Base for males and females respectively. For the purposes of those tables, adjustments were made for ages below 65 and above 91 in the case of males and below 67 and above 95 in the

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

Table 1. Chi-squared test on pensioners graduations

Type	Gender	Formula	χ^2	Degrees of freedom	$p(\chi^2)$
Normal or late retirement					
	Males	GM(1,3)	54.72	43	0.1085
	Females	GM(1,3)	63.28	39	0.0083
Early retirement					
	Males	GM(2,2)	80.43	51	0.0054
	Females	GM(2,2)	53.68	41	0.0886

case of females. Those adjustments are ignored for the purposes of this paper. The results of the chi-squared test on those graduations were as shown in Table 1.

5.1.2 It is suggested by Forfar, McCutcheon & Wilkie (1988) that a graduation is satisfactory if the value of $p(\chi^2)$ is greater than 0.05. On this criterion, only the normal or late retirement males and the early retirement females are satisfactory. Even those graduations are not ideal. The others are clearly unsatisfactory.

5.1.3 In commenting on the graduation of the normal or late retirement males, Forfar, McCutcheon & Wilkie (1988) stated that they had:

> "no reason to suppose that a large number of duplicates existed ..., and no information from which variance ratios could be derived, so it was assumed that each observation represented one 'life'."

5.1.4 For the purposes of illustration, the methods outlined in this paper have been applied to the graduations referred to in ¶5.1.1, which have been interpreted as graduations of annual average forces of mortality.

5.2 *Parameters*

5.2.1 It was assumed that there were no duplicates, so equation (15) was used for $F(y)$.

5.2.2 It was found that the values $\varepsilon_1 = 0.05$ and $\varepsilon_2 = 1e^{-10}$, which are required for the algorithms in Appendices A and B respectively, gave sufficient accuracy without resulting in excessively long running times. (Where it is more convenient, numbers are expressed in scientific notation, i.e. in the form Xe^Y instead of $Z = X.10^Y$.)

5.2.3 For each graduation the results depend on b (the power of the smoothness function) and on η_1 and η_2 (the weightings of the components of the smoothness function), as defined in ¶3.3, as well as on the Lagrange multiplier λ. For each graduation the program was run for the combinations of b, η_1 and η_2 and the ranges of λ (from $\lambda_{\min}$ to $\lambda_{\max}$) shown in Table 2. In general, the ranges of values of λ were selected so that the ranges of values of $p(\chi^2_\nu)$ corresponded approximately to equation (68). Exceptions were made where the smoothness deteriorated to unacceptable levels; in those cases the ranges of values of λ were further restricted. As explained in ¶3.5, the values

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

Table 2. Parameter combinations and ranges of λ

	Males			Females		
Combination number	1	2	3	1	2	3
b	2	4	2	2	4	2
Normal or late retirement						
η_1	0	0	0.19	0	0	0.21
η_2	1	1	0.81	1	1	0.79
λ_{min}	$6.89e^4$	$2.5e^{10}$	$7e^4$	$1e^2$	$6e^4$	$1e^2$
λ_{max}	$1e^7$	$1e^{15}$	$1e^7$	$1e^7$	$1e^9$	$1e^5$
Early retirement						
η_1	0	0	0.91	0	0	0.99
η_2	1	1	0.09	1	1	0.01
λ_{min}	$1e^4$	$1e^8$	$1e^5$	$1e^3$	$5.85e^6$	$8e^4$
λ_{max}	$1e^6$	$2e^{11}$	$1e^7$	$1e^5$	$1e^{10}$	$1e^7$

of $z(1)$ and $z(2)$ (the orders of differencing of the smoothness function) were taken as respectively equal to r and s (the type parameters of the parametric graduation formula defined in ¶3.1).

5.2.4 In Forfar, McCutcheon & Wilkie (1988) the GM and LGM formulae are expressed in terms of orthogonal polynomials. Furthermore, in that paper, age-scaling was applied so that the argument of the polynomials was:

$$\frac{\text{age} - 70}{50}.$$

In order to apply the methods of this paper, it is necessary to convert the coefficients a_i and b_i defined by Forfar, McCutcheon & Wilkie (1988, pp61-62) for the age-scaled orthogonal polynomials, to the corresponding coefficients α_i for the ordinary polynomials t_{1x} and t_{2x} used in Section 3 of this paper. The formulae for these conversions are as follows:

(1) for a first-order polynomial: $\alpha_1 = a_0$;
(2) for a second-order polynomial: $\alpha_1 = a_0 - 1.4a_1$; $\alpha_2 = 0.02a_1$; and
(3) for a third-order polynomial: $\alpha_1 = a_0 - 1.4a_1 + 2.92a_2$; $\alpha_2 = 0.02a_1 - 0.112a_2$; $\alpha_3 = 0.0008a_2$.

Similar relationships apply between α_{r+j} and b_j. The resulting coefficients are determined as shown in Table 3.

5.3 *Tests*

In the program a number of tests are applied to the original graduations and to the regraduations. The tests applied are the chi-squared test, the signs test and the runs test. All tests are made after grouping as described in ¶2.8. To test the significance of the number of runs, the formula given by Forfar, McCutcheon & Wilkie (1988, p44) for the probability that the number of runs (NR) is equal to r are used.

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

Table 3. Determination of coefficients

	Normal or late retirement		Early retirement	
	Males	Females	Males	Females
$100a_0$	0.557291	0.662810	−14.141218	−2.005646
$100a_1$			−23.423078	−5.687258
b_0	−4.993529	−6.473887	−1.659785	−3.108251
b_1	5.882482	8.069982	2.271733	3.838825
b_2	−1.668855	−2.174915		
α_1	0.00557291	0.0066281	0.186510912	0.059565152
α_2	−18.1020604	−24.1226136	−0.0046846156	−0.0011374516
α_3	0.3045614	0.40499012	−4.8402112	−8.482606
α_4	−0.001335084	−0.001739932	0.04543466	0.0767765

5.4 *Results*

5.4.1 It should be noted that values of $S(v)$ cannot be compared between different combinations of b, η_1 and η_2. For each such combination, the values of $F(v)$ for the respective graduations are shown as a function of $s = S(v)$ in Figure 1. The curves shown in that figure are analogous to the efficient frontiers of portfolio theory. For a particular value of the smoothness coefficient, each point on a curve represents the regraduation with the greatest likelihood (i.e. with the lowest value of $F(v)$). It may be noted that, for all parameter combinations, $F(\cdot)$ is more sensitive to $s = S(\cdot)$ for males early retirement than for the other data sets. Females show relatively low sensitivity, particularly for normal retirement with parameter combination 3. The points shown as 'selected regraduation' are explained below.

5.4.2 Consideration now needs to be given to the selection of a particular regraduation for each data set. This process is illustrated below with reference to males early retirement.

5.4.3 The results of the regraduation of males early retirement with parameter combination 1 are shown in Table 4 for the extreme values $\lambda_{\min}$ and $\lambda_{\max}$ used for Figure 1 and for three geometrically interpolated values.

5.4.4 The values of $\bar{\mu}_x$ for $\lambda = 1e^4$ are shown in Figure 2. The upper frame shows the full range of ages and the lower frame shows the central range from 55 to 85. Each frame shows, on a scale appropriate for its age domain, the observed rates $\tilde{v}_x$, the original graduation u_x and the regraduation v_x, as well as the natural logarithm of the ratio of v_x to u_x. The turning-points in the curve of the logarithm of the ratio indicate the positions of extremes in the departure of the regraduation from the original graduation. The curve representing the regraduation may be examined, particularly at those points, to consider whether the regraduation is sufficiently smooth. The dots and dashes representing the observed values may help to explain the divergence of the regraduation from the original graduation. It appears that there is room for improvement in the smoothness of the regraduation, particularly in the regions of ages 55 and 68 where there are downward kinks, as well as the region of age 45 where there is a marked upward kink. The kinks at ages 45 and 55 are particularly problematic

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

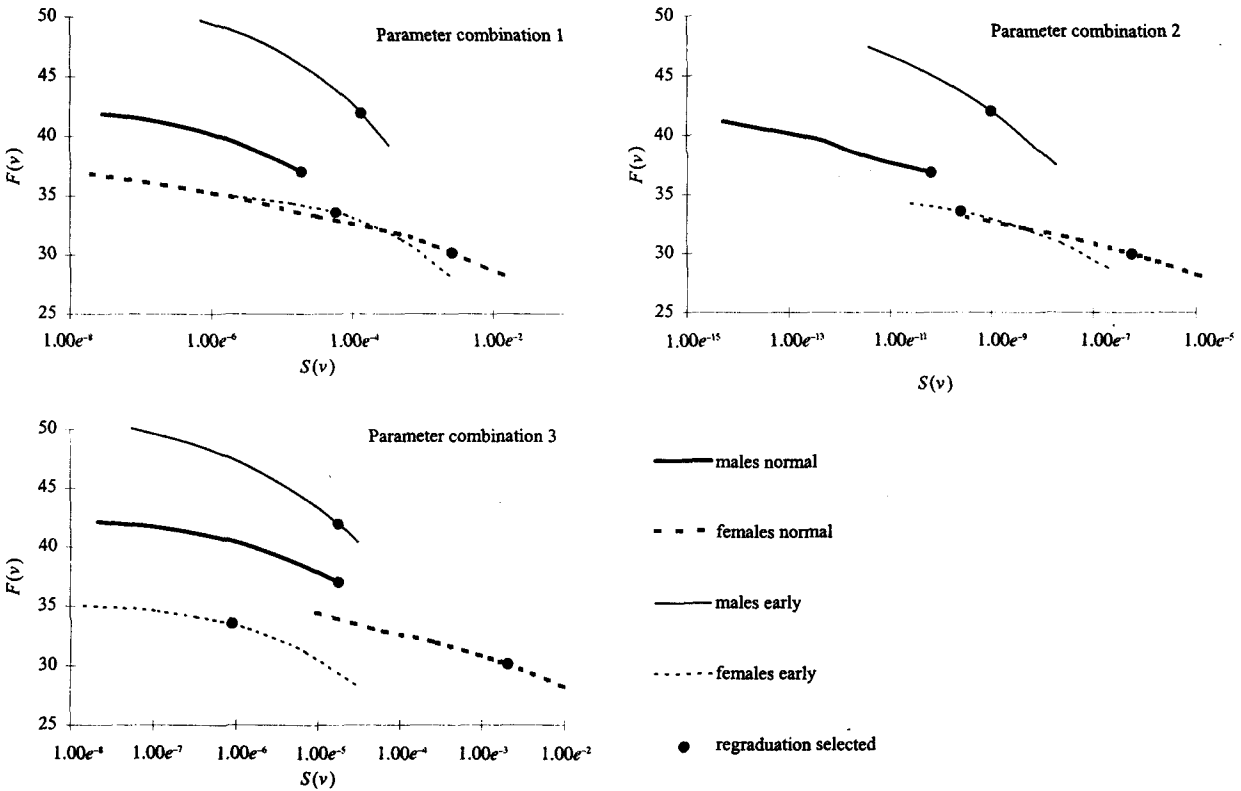


Figure 1. $F(v)$ and $S(v)$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

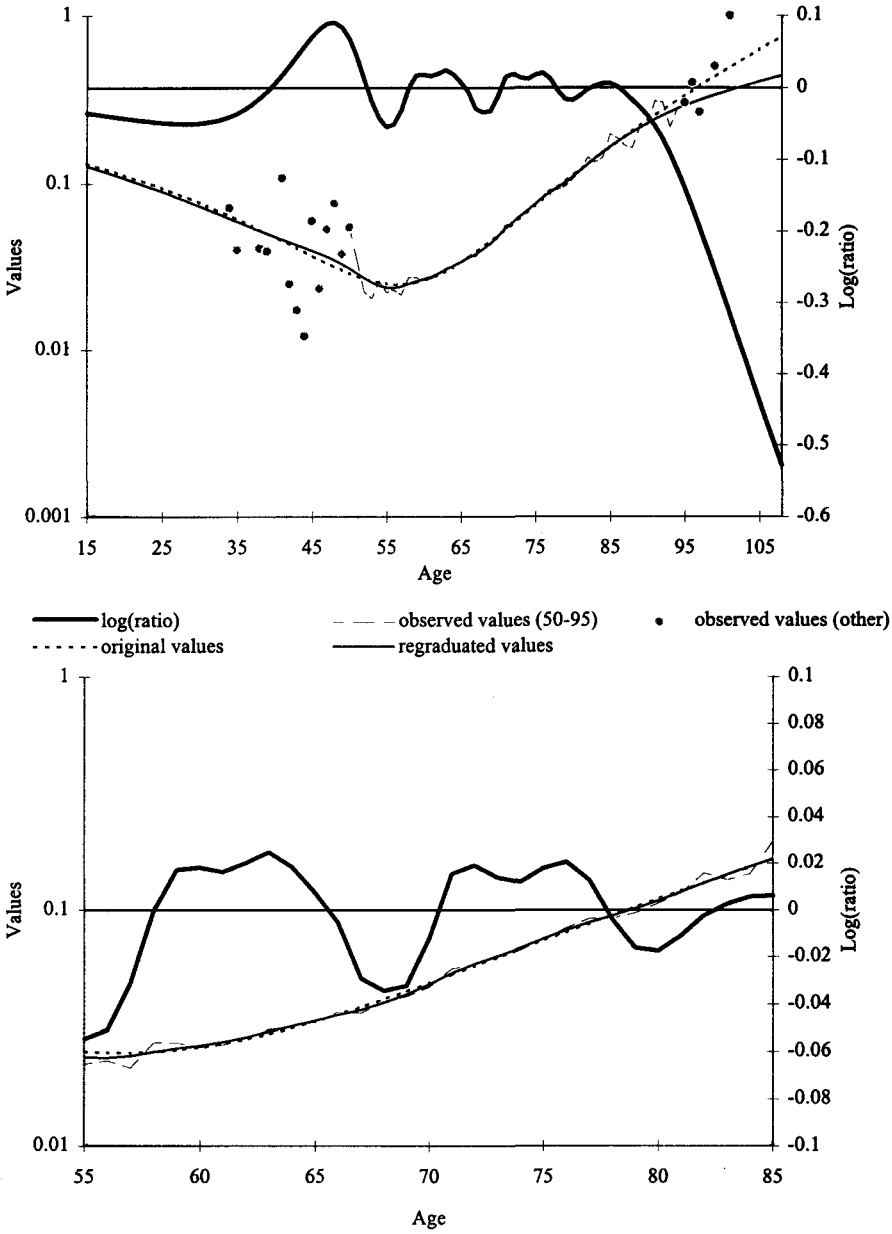


Figure 2. Males early retirement; parameter combination 1; $\lambda = 1e^4$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

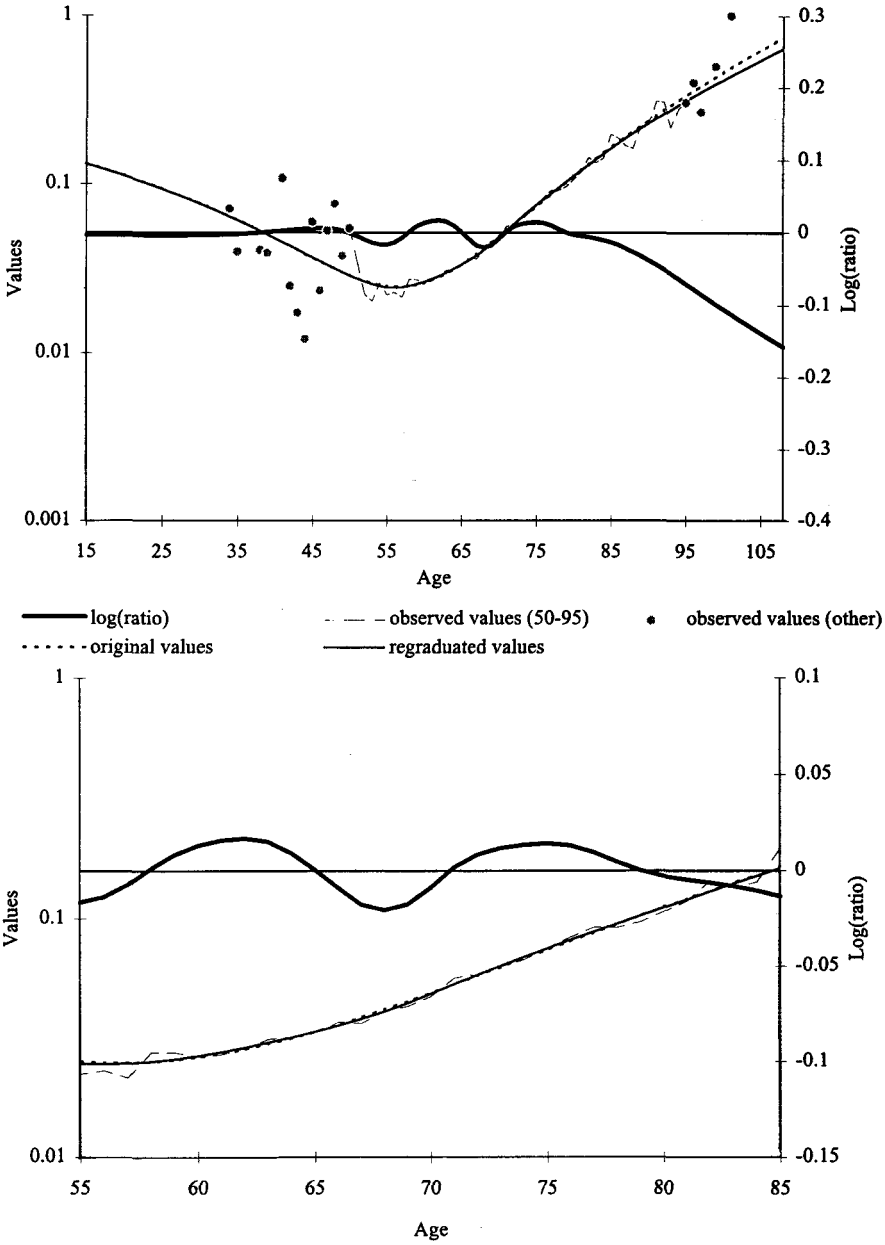


Figure 3. Males early retirement; parameter combination 1; $\lambda = 1e^5$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

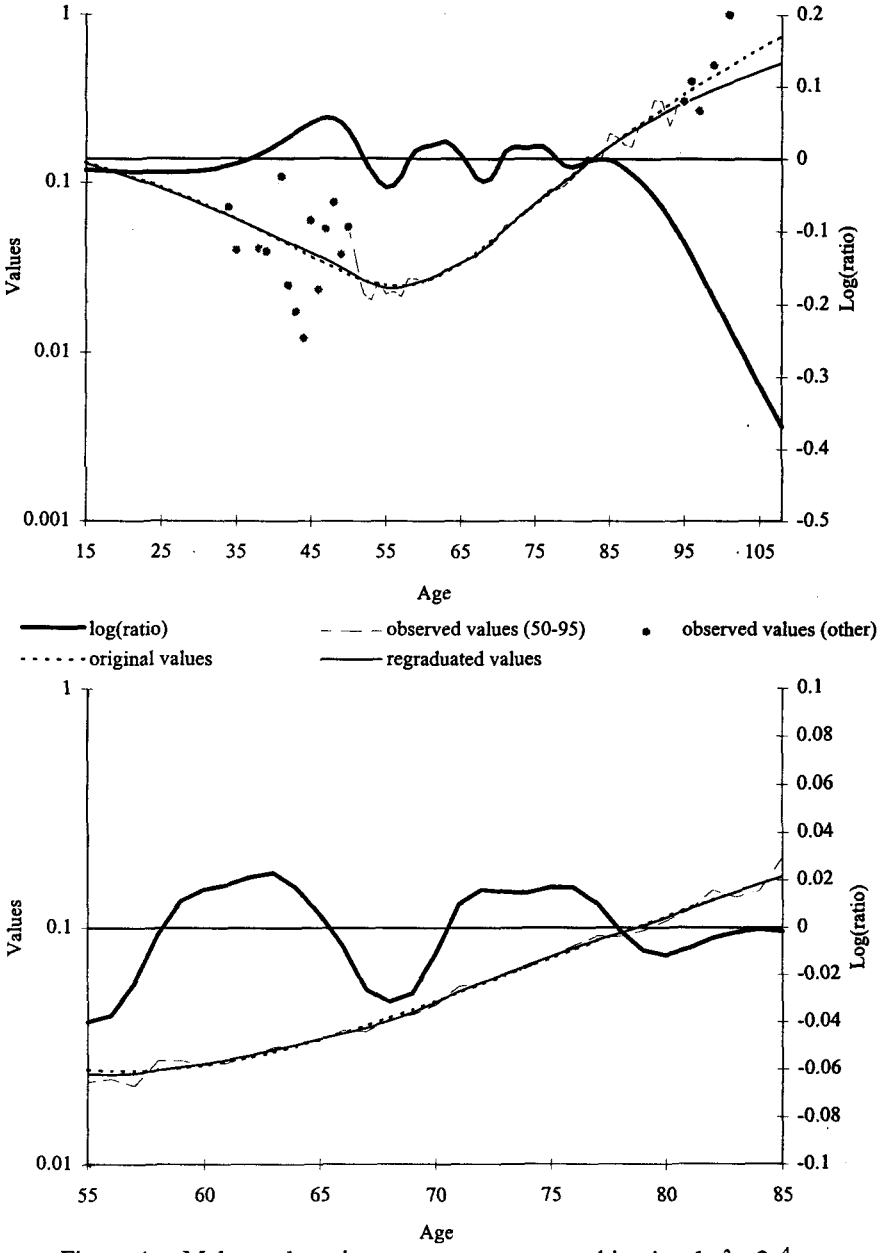


Figure 4. Males early retirement; parameter combination 1; $\lambda=2e^4$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

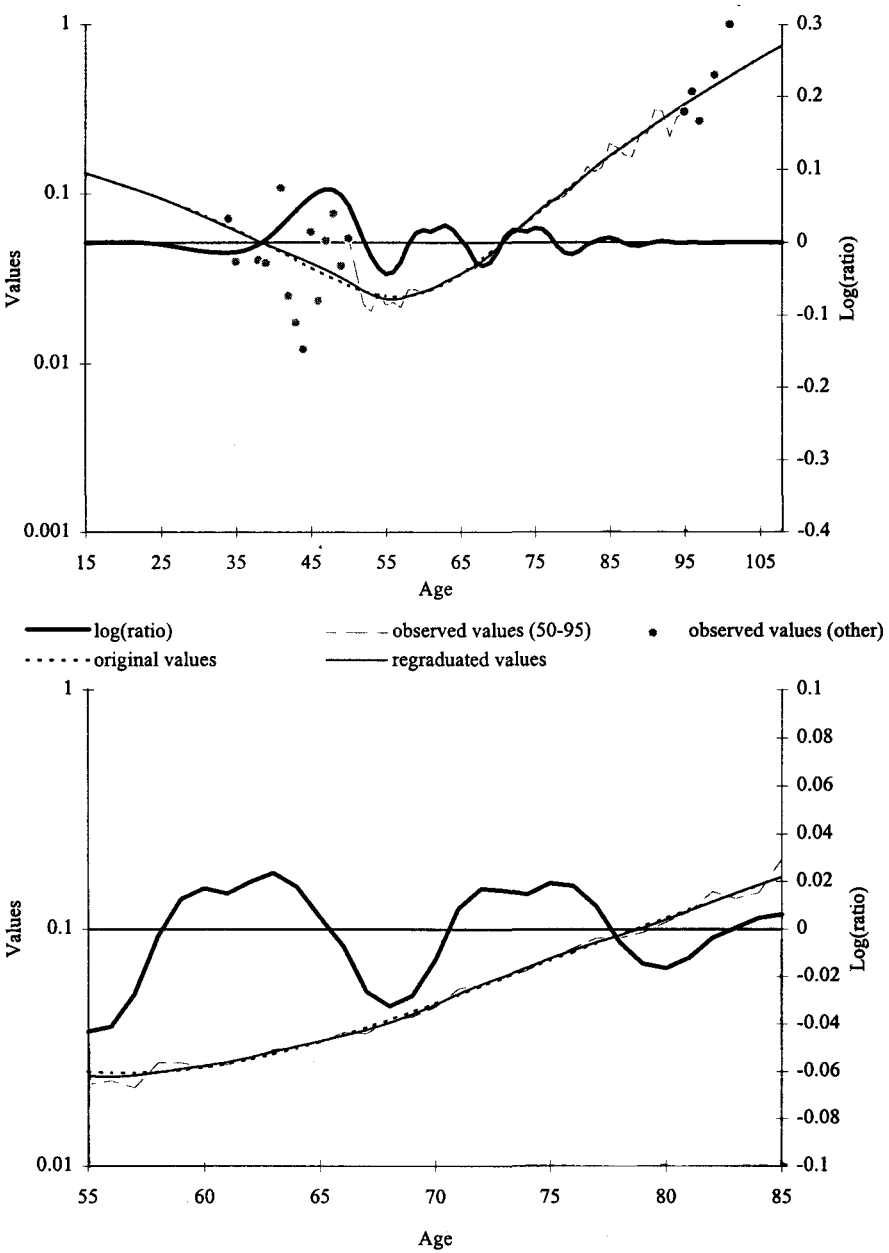


Figure 5. Males early retirement; parameter combination 2; $\lambda = 1.345e^{9}$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

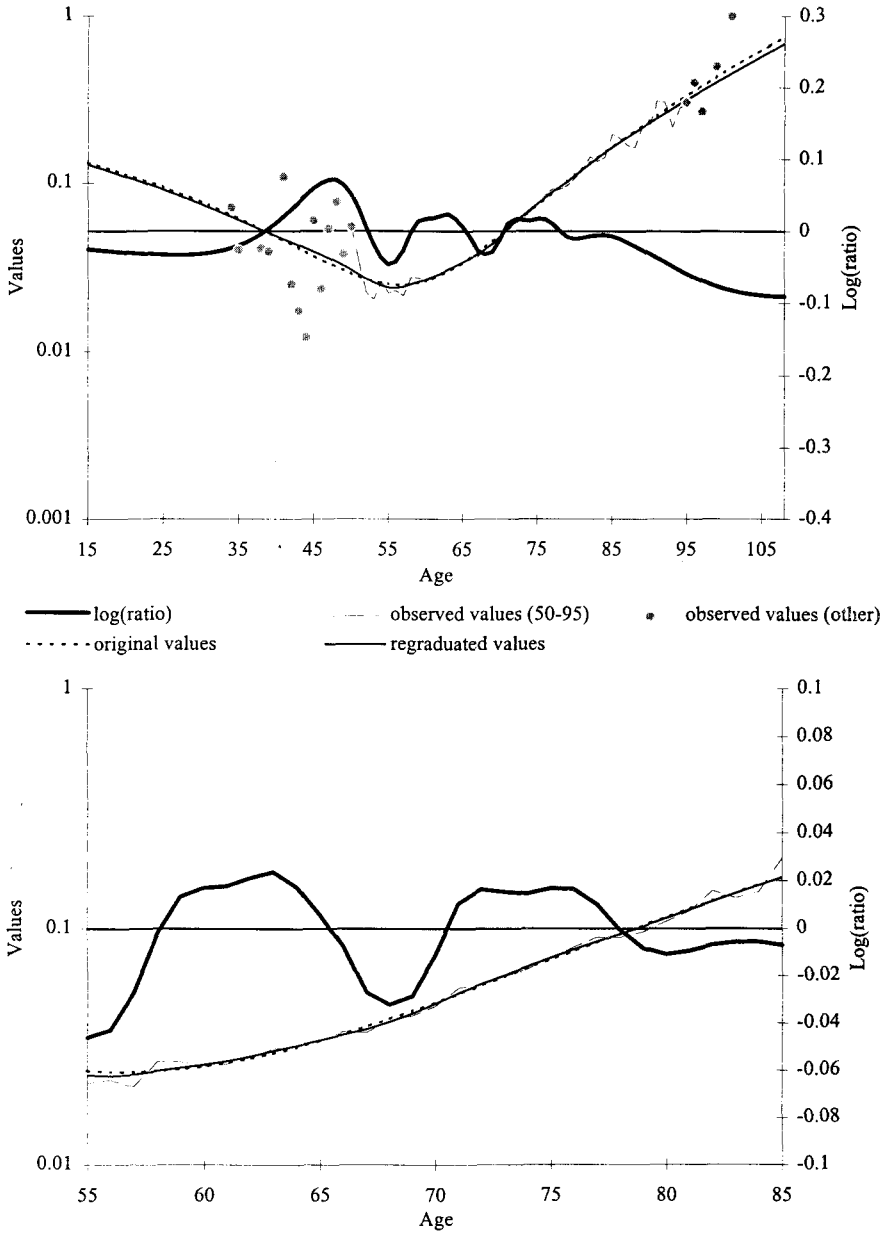


Figure 6. Males early retirement; parameter combination 3; $\lambda = 1.492e^5$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

Table 4. Results of regraduations of males early retirement parameter combination 1

λ	$S(v)$	$F(v)$	$p(\chi^2_v)$
$1.00e^4$	$3.37e^{-4}$	39.1	0.293
$3.16e^4$	$7.02e^{-5}$	43.4	0.089
$1.00e^5$	$1.57e^{-5}$	46.2	0.034
$3.16e^5$	$3.58e^{-6}$	48.2	0.017
$1.00e^6$	$6.83e^{-7}$	49.7	0.010
		$F(u)$	$p(\chi^2_u)$
Original graduation		51.3	0.005

because they occur in the vicinity of the turning-point of the graduation, which tends to exaggerate them. It is therefore pointless to consider regraduations with $p(\chi^2_v) > 0.293$. However, smoother regraduations with $p(\chi^2_v) < 0.293$, i.e. for $\lambda > 1e^4$, may be considered.

5.4.5 The values of $\bar{\mu}_x$ for $\lambda = 1e^5$ are shown in Figure 3. For ease of comparison the scales are the same as for Figure 2. This regraduation is very smooth; although slight fluctuations are evident in the ratios, they are not reflected by noticeable kinks in the regraduated values. An intermediate value of λ may therefore be considered.

5.4.6 The results of a regraduation of $\bar{\mu}_x$ for $\lambda = 2e^4$ are as follows:

$$S(v) = 1.31e^{-4}$$
$$F(v) = 41.9$$
$$p(\chi^2_v) = 0.141.$$

The values of $\bar{\mu}_x$ on this regraduation are shown in Figure 4. The smoothness of this regraduation was considered reasonable; the turning-points in the ratio do not appear to produce unacceptable kinks on either of the scales presented. The point in Figure 1 represented by these parameters is shown as the selected regraduation for parameter combination 1.

5.4.7 The next step is to consider alternative parameter combinations. For combinations 2 and 3, regraduations were found that gave approximately the same value of $F(x)$ as for the above regraduation. The results are shown in Table 5. The values of $\bar{\mu}_x$ on these regraduations are shown in Figures 5 and 6.

5.4.8 Although there is very little difference between the three regraduations, except at the highest ages, it appears from Figures 4 to 6 that the smoothest regraduation of the three is for parameter combination 1. This regraduation was therefore adopted.

5.4.9 Similar processes were followed for the other data sets, except that, for males normal or late retirement, the three regraduations that gave $p(\chi^2_v) = 0.5$ were sufficiently smooth; the smoothest of those three regraduations was adopted.

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

Table 5. Results of regraduations of males early retirement

Parameter combination	λ	$S(v)$	$F(v)$	$p(\chi^2_v)$
1	$2.000e^4$	$1.31e^{-4}$	41.9	.141
2	$1.345e^9$	$9.74e^{-10}$	42.0	.161
3	$1.492e^5$	$1.76e^{-5}$	41.9	.153

The values of $\bar{\mu}_x$ on the regraduations adopted for males and females normal or late retirement and for females early retirement are shown in Figures 7 to 9. The results of the adopted regraduations, including the full range of tests, are shown in Table 6. In that table the adopted values of λ are denoted by λ_{obj}.

5.4.10 In each case the likelihood was considerably enhanced, with the exception of females early retirement, where the enhancement was relatively modest. The results of the χ^2 tests are considerably improved, although, in this context, the remarks made in ¶4.15 should be borne in mind. The results of the signs tests show no significant underfitting or overfitting. The results of the runs tests are greatly improved for early retirement. In the case of males early retirement, the result is no longer significantly low. For normal or late retirement the runs tests seem to suggest overfitting, particularly for females. However, this is not borne out by the χ^2 test, and was, in any case, evident, even in the parametric graduations.

5.4.11 It is of interest that the ratios of the regraduated values to the original graduations show marked cyclical behaviour in the cases of males and females early retirement, as may be seen from Figures 4 and 9. The question whether this feature is ephemeral or related to age or age cohorts — or an artificial product of the enhancement process — is a matter for further research.

5.4.12 The shape of the early retirement curves is unusual, presumably due to the prevalence of substandard lives amongst the exposed to risk at younger ages. However, it is noted that the regraduations did not result in any serious deterioration of the shape of the curves

5.5 *The Normal Approximation*

5.5.1 In the presence of duplicates it is necessary to use the normal approximation. In the absence of duplicates the question arises whether it is better to use the true likelihood function. For each of the adopted regraduations, an alternative regraduation was made using the normal approximation. For each alternative regraduation approximately the same smoothness constraint was used as for the corresponding regraduation with the true likelihood function. The results are shown in Table 7.

5.5.2 The normal approximation shows some deterioration of the results of the tests. In the case of the χ^2 test, the deterioration is marginal. In the case of the signs test the result becomes significantly low for males early retirement; for other data sets there is no deterioration. In the case of the runs test the result again deteriorates for males early retirement. Inspection of the regraduated values

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

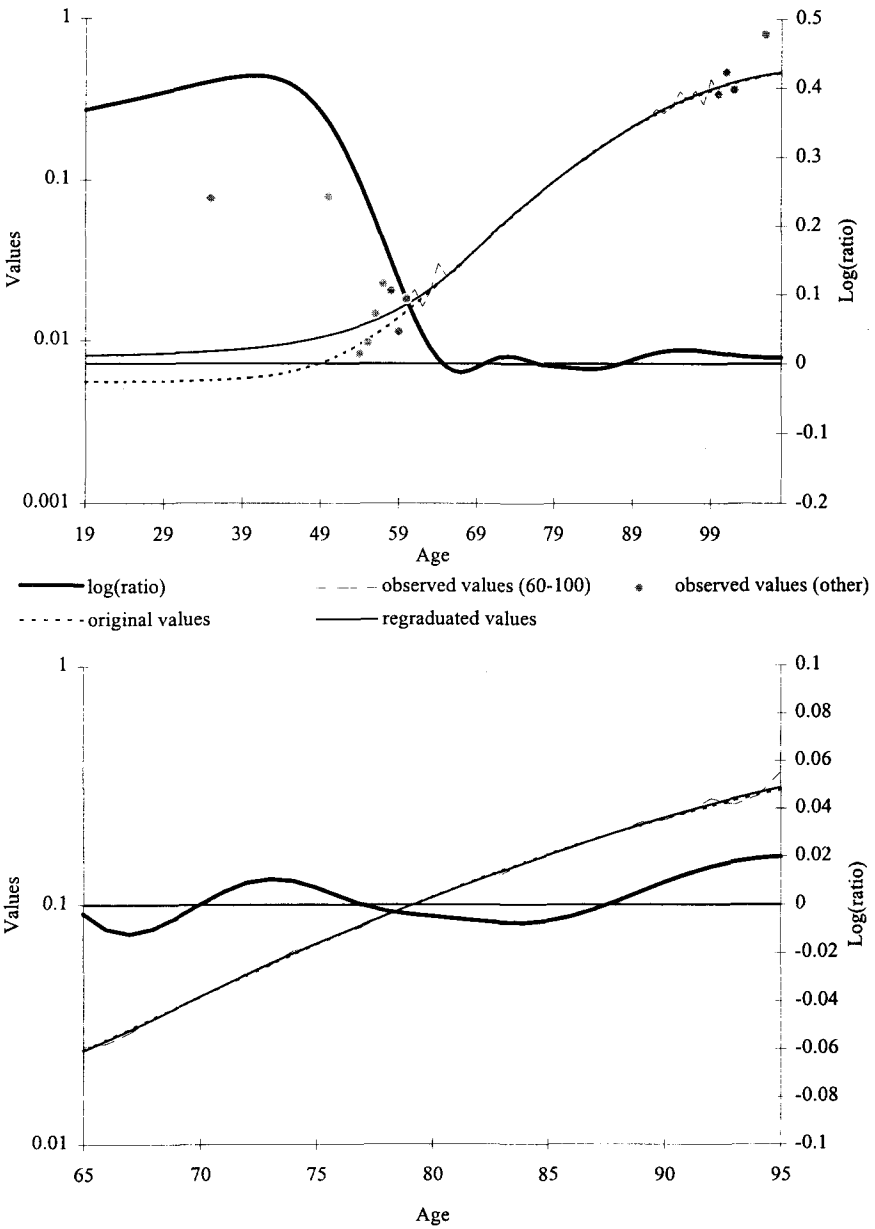


Figure 7. Males normal or late retirement; parameter combination 3; $\lambda = 7e^4$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

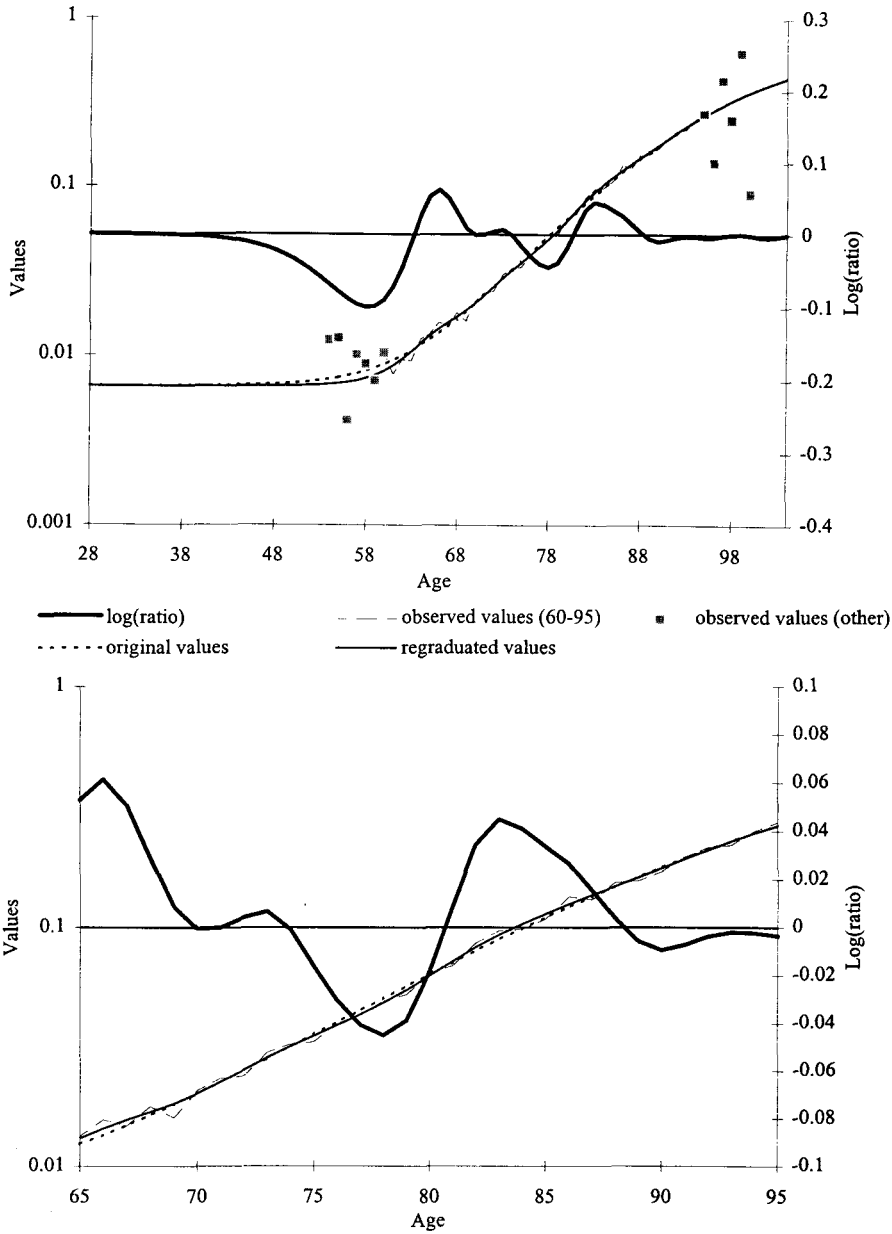


Figure 8. Females normal or late retirement; parameter combination 1; $\lambda=4e^2$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

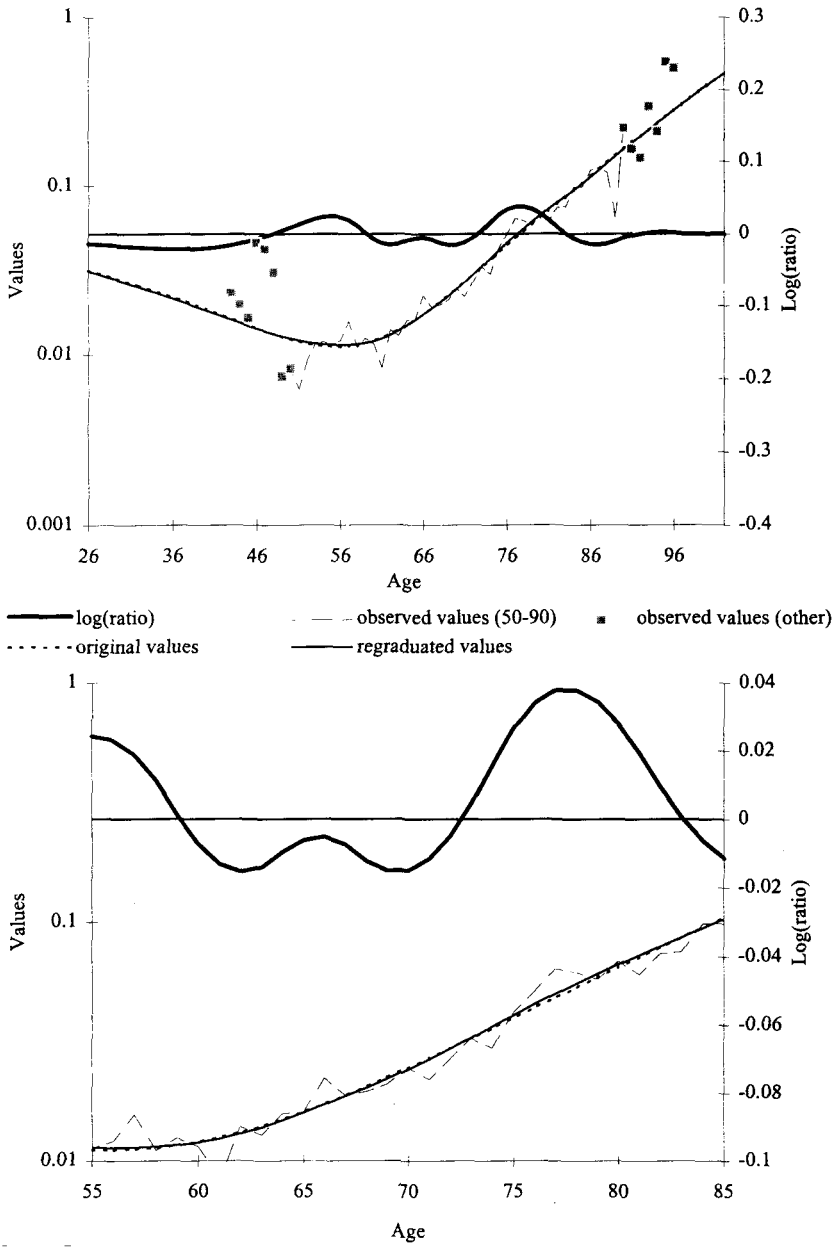


Figure 9. Females early retirement; parameter combination 2; $\lambda = 1.56e^{9}$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

Table 6. Results of tests

	Normal or late retirement				Early retirement			
	Males		Females		Males		Females	
	Original	Enhanced	Original	Enhanced	Original	Enhanced	Original	Enhanced
Parameter combination		3		1		1		2
b		2		2		2		4
η_1		0.19		0		0		0
η_2		0.81		1		1		1
λ_{obj}		$7e^4$		$4e^2$		$2e^4$		$1.56e^9$
s	0	$1.79e^{-5}$	0	$2.58e^{-3}$	0	$1.31e^{-4}$	0	$3.14e^{-10}$
$F(y)$	42.7	37.0	37.3	30.1	51.3	41.9	35.3	33.5
χ^2_ν	54.7	42.2	63.3	49.0	80.4	61.9	53.7	49.8
degrees of freedom	43	39	51	41				
$p(\chi^2_\nu)$	0.109	0.505	0.008	0.132	0.005	0.141	0.089	0.162
n_1	23	24	21	24	25	24	23	22
n_2	24	23	22	19	30	31	22	23
p(signs)	0.500	0.615	0.500	0.820	0.295	0.209	0.617	0.500
runs	29	29	27	29	21	27	25	27
p(runs)	0.930	0.930	0.939	0.990	0.031	0.438	0.727	0.886

Table 7. Comparison of results based on the normal approximation with those based on the true likelihood function

	Normal or late retirement				Early retirement			
	Males		Females		Males		Females	
	True	Normal	True	Normal	True	Normal	True	Normal
s	$1.79e^{-5}$	$1.76e^{-5}$	$2.58e^{-3}$	$2.58e^{-3}$	$1.31e^{-4}$	$1.34e^{-4}$	$3.14e^{-10}$	$3.13e^{-10}$
$F(y)$	37.0	43.2	30.1	49.2	41.9	62.0	33.5	49.7
χ^2_ν	42.2	43.2	49.0	49.2	61.9	62.0	49.8	49.7
degrees of freedom	43	39	51	41				
$p(\chi^2_\nu)$	0.505	0.462	0.132	0.126	0.141	0.139	0.162	0.166
n_1	24	24	24	24	24	21	22	22
n_2	23	23	19	19	31	34	23	23
p(signs)	0.615	0.615	0.820	0.820	0.209	0.052	0.500	0.500
runs	29	29	29	29	27	23	27	29
p(runs)	0.930	0.930	0.990	0.990	0.438	0.160	0.886	0.966
$\bar{\mu}$								
age								
50	0.0108580	0.0096975	0.0066664	0.0067855	0.0300424	0.0302879	0.0124232	0.0126172
60	0.0167565	0.0165614	0.0081307	0.0082297	0.0265956	0.0265972	0.0119643	0.0119520
70	0.0415615	0.0415614	0.0201974	0.0202417	0.0482200	0.0482130	0.0242569	0.0241968
80	0.1071634	0.1071424	0.0627890	0.0629791	0.1099965	0.1097938	0.0666938	0.0664004
90	0.2295280	0.2285634	0.1758119	0.1766914	0.2274011	0.2362511	0.1637192	0.1644684
100	0.3805119	0.3747499	0.3660673	0.3660633	0.3728196	0.4610111	0.3926723	0.3928511

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

shows very little difference between the true likelihood function and the normal approximation, except at the upper extreme of the age range for males early retirement.

5.5.3 In general, the normal approximation appears to be satisfactory, but the evidence of males early retirement suggests that, if possible, the true likelihood function should be used.

6. CONCLUSION

The method outlined in this paper may be used to enhance the likelihood, and thus the fidelity, of a parametric graduation. The method is effectively based on a trade-off between fidelity and smoothness. The specification of the fidelity criterion is objective, whereas that of the smoothness criterion may include subjective elements (in particular the parameter b). In some cases the fidelity criterion can be improved to the point at which $p(\chi^2)=0.5$. In such cases that point provides an objective selection of a final regraduation. In other cases the selection process is subjective, in that it depends on a visual comparison of the regraduated values with the original graduation. Although the method imposes additional complexity on a parametric graduation, it provides a workable way forward, not only when parametric graduations result in significant departures from fidelity, but also when it is desired to improve the fidelity of a parametric graduation subject to reasonable smoothness constraints.

ACKNOWLEDGEMENTS

The author acknowledges with thanks the help of Mrs J.V. Evans and Mr G.J. Farmar with regard to various aspects of this paper, and the very helpful comments and suggestions made by the scrutineers.

REFERENCES

CHAN, F.Y., CHAN, L.K. & YU, M.H. (1984). A generalization of Whittaker-Henderson graduation. *T.S.A.* **36**, 183-211.

CMI COMMITTEE (1988). The graduation of the 1979-82 mortality experiences. *C.M.I.R.* **9**, 1-102.

CMI COMMITTEE (1990). Standard tables of mortality based on the 1979-82 experiences. *C.M.I.R.* **10**, 1-138.

FORFAR, D.O., MCCUTCHEON, J.J. & WILKIE, A.D. (1988). On graduation by mathematical formula. *J.I.A.* **115**, 1-149.

LONDON, D. (1985). *Graduation: the revision of estimates*, ACTEX, Winsted.

WALSH, G.R. (1975). *Methods of optimization*, Wiley, London.

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

APPENDIX A

ALGORITHM FOR THE DFP FORMULA

A.1 The purpose of this algorithm, which is restated from Walsh (1975, pp110-117), is to obtain y^* such that $DM(y^*)=0$ and $D(DM(y^*))^{\mathsf{T}}$ is a positive-definite matrix.

A.2 For $k = 1, \ldots :$

A.2.1
$$d_k = -H_k DM(y_k) \tag{A1}$$

with:

$$H_k = 10^{-8} I \text{ for } k=1 \tag{A2}$$

and H_k as defined below for $k > 1$.

The usual specification of this algorithm starts with $H_1 = I$. However, in this application, that would result in inaccuracies in equation (A9), where each of the expressions on the right-hand side would be of the same order of magnitude, but the absolute values of the elements of H_k would decrease substantially as k increases.

(Note: d_k defines the direction of search from the current point y_k.)

A.2.2 Perform a linear search, such as Davidon's cubic interpolation method, which is set out in Appendix B, to find β_k^*, i.e. the value of β_k that minimises $M(y_k+\beta_k d_k)$.

A.2.3
$$p_k = \beta_k d_k \tag{A3}$$

$$y_{k+1} = y_k + p_k \quad \text{giving the next current point} \tag{A4}$$

$$q_k = DM(y_{k+1}) - DM(y_k) \tag{A5}$$

$$H_k^{(1)} = \frac{p_k p_k^{\mathsf{T}}}{p_k^{\mathsf{T}} q_k} \tag{A6}$$

$$J_k = H_k q_k \tag{A7}$$

$$H_k^{(2)} = \frac{J_k J_k^{\mathsf{T}}}{q_k^{\mathsf{T}} J_k} \tag{A8}$$

$$H_{k+1} = H_k + H_k^{(1)} - H_k^{(2)}. \tag{A9}$$

A.2.4 If $M(y_{k-n}) - M(y_k)$ is less than ε_1 and more than $2n$ iterations have been

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

made, then $y^* = y_{k+1}$. To avoid the possibility of a closed loop, a limit of $5n$ was placed on the number of iterations, but with the parameters as specified, the algorithm converged before that limit was reached.

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

APPENDIX B

ALGORITHM FOR MODIFIED DAVIDON'S CUBIC INTERPOLATION METHOD

B.1 The purpose of this algorithm, which is restated from Walsh (1975, pp97-102) is to obtain β^* such that $M(y+\beta^* d) \leq M(y+\beta d)$ for $\beta \neq \beta^*$.

B.2 $$a = 0 \tag{B1}$$

$$b = e(y). \tag{B2}$$

B.3 For $k = 1,2,\ldots$:

B.3.1 $$M_a = M(y + ad) \tag{B3}$$

$$M_a' = (DM(y+ad))^{\top} d \tag{B4}$$

$$M_b = M(y + bd) \tag{B5}$$

$$M_b' = (DM(y+bd))^{\top} d. \tag{B6}$$

B.3.2 If $M_b' > 0$ or $M_b > M_a$, then:

B.3.2.1 $$z = \frac{3}{b-a}(M_a - M_b) + M_a' + M_b' \tag{B7}$$

$$w = \sqrt{z^2 - M_a' M_b'} \tag{B8}$$

$$c = a + \left(1 - \frac{M_b' + w - z}{M_b' - M_a' + 2w}\right)(b-a) \tag{B9}$$

$$M_c' = (DM(y+cd))^{\top} d \tag{B10}$$

$$\varepsilon = \min(c-a,\ b-c). \tag{B11}$$

If $M_c' < 0$ then:

$$a = c \tag{B12}$$

otherwise:

$$b = c \tag{B13}$$

otherwise:

B.3.2.2 $$b = e(y + bd) \tag{B14}$$

$$c = b. \tag{B15}$$

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press

B.3.3 If ϵ is less than a specified constant ϵ_2, or if a specified maximum number of iterations have been completed, then:

$$\beta^* = a,\ b \text{ or } c \text{ according as } M_a,\ M_b \text{ or } M_c \text{ is the least.} \qquad \text{(B16)}$$

For the purposes of ¶B.3.3 it was found that, in the application of the algorithm, the optimal maximum number of iterations was 4.

https://doi.org/10.1017/S1357321700000428 Published online by Cambridge University Press