JIA 120 (1993) 153-170

J.I.A. **120**, I, 153–170

GRADUATION BY DYNAMIC REGRESSION METHODS

BY R. J. VERRALL, M.A., M.Sc., Ph.D.

(of The City University, London)

ABSTRACT

This paper extends the theory of graduation by parametric formulae to include dynamic estimation methods. This is an application of the Kalman filter and allows the parameters of the curve fitted to vary with age. The amount of variation is determined by the amount of smoothing required, and the method can be regarded as a combination of curve fitting and sequential smoothing, each of which has been used separately for performing graduations. In practice, a dynamic straight line can always be used for the graduation and the method has a sensible logical interpretation.

KEYWORDS

Graduation; Kalman Filter; Linear Models; Mortality

1. INTRODUCTION

The actuarial art of graduation often involves the application of linear and non-linear models. A summary of these models within the framework of generalised linear models is given by Renshaw (1991) and a comprehensive overview of current actuarial practice is given by Forfar *et al.* (1988). Both of these papers concentrate on the use of parametric modelling methods to obtain a reasonably close fit to actual experience, giving due weight to model parsimony and smoothness. In doing this, it is usual to fit a curve (a straight line, cubic, quartic, etc.) which is assumed to apply to the data over the whole range of ages considered. It is sometimes the case that a straight line produces an adequate fit (to the transformations of the death rate considered in this paper), but greater flexibility can be obtained by using higher order polynomials, non-linear regression functions or, within the context of generalised linear models, non-canonical link functions.

This paper proposes a different method of obtaining greater flexibility than can be obtained by simply fitting a straight line. The technique proposed has greater heuristic appeal than fitting higher order polynomials, and the form of the flexibility used appears to be very natural in the context of graduation. It involves a sophisticated smoothing procedure (retaining the parametric form of the regression), which can also be related to more basic methods of smoothing which seem natural for graduation, and have been used in the past. When, for example, a quartic curve is used in a graduation, it is natural to question whether such a specific curve has any meaning for the statistics graduated, or whether it is more sensible to apply a combination of curve fitting (of low order) and sequential smoothing. A straight line fit and a smoothing procedure both clearly have straightforward interpretations in graduation. The aim of this paper is to combine these two approaches, and obtain a sequential smoothing procedure which can be used in combination with a straight line regression model (applied to the transformation of the death statistics).

Two examples will be used, which have been extracted from the CMI Committee (1976) (C.M.I. Report No. 2) and Forfar *et al.* (1988). These examples cover the graduation of both the probability of death q_x, and the force of mortality μ_x, over a range of ages which are indexed by x.

Section 2 reviews graduation theory and the use of generalised linear models. Section 3 extends the theory to dynamic generalised linear models and Section 4 contains the examples.

2. Graduation by Generalised Linear Models

In general terms, the aim of graduation is to produce a smooth table of death statistics from a set of raw data. For this type of graduation, the data comprise the exposures R_x, and the actual numbers of deaths A_x, for each age x. The exposure at age x, R_x, is the initial exposed to risk when graduating q_x and the central exposed to risk when graduating μ_x. These definitions are regarded as conventions throughout this paper.

Two quantities are of interest: the probability of death at each age q_x, and the force of mortality at each age μ_x. These quantities can be estimated separately at each age, the maximum likelihood estimates for each being:

$$\hat{q}_x = \frac{A_x}{R_x} \qquad \hat{\mu}_x = \frac{A_x}{R_x} \tag{2.1}$$

where R_x is interpreted as the initial and the central exposed to risk in each of these equations, respectively.

For q_x, the observed number of deaths A_x, is assumed to be binomially distributed with parameters R_x and q_x. For μ_x, the distribution of A_x is assumed to be Poisson with parameter $R_x\mu_x$. Section 2 of Forfar *et al.* (1988) derives these estimates, which are, of course, trivial examples of maximum likelihood theory.

Thus A_x/R_x can be thought of as the crude estimator of q_x and μ_x when each age is analysed individually.

Clearly it would be unsatisfactory to present a table consisting of these crude estimates which would probably imply sharp changes in the death rates from age to age. If this were done some important information would have been ignored; this is that the actual values of q_x and μ_x should change gradually and smoothly with age x. This leads to the use of smoothing methods to produce a continuous table of death rates, and, in particular, the use of regression methods. It is clear that linear regression cannot be applied, since the data are not normally distributed. Generalised linear models are designed to handle non-normal data, and can be used for graduation. The standard text on the theory of generalised linear models is McCullagh & Nelder (1989), and Renshaw (1991) contains greater details on the application to graduation. A sketch of the general theory is now given.

The likelihood, which will involve some sort of parametrised curve, is based on either the binomial distribution (for q_x) or the Poisson distribution (for μ_x). Forfar *et al.* (1988) approach the maximum likelihood theory via what they term the 'Gompertz-Makeham formula of type (r,s)' and the 'Logit Gompertz-Makeham formula of type (r,s)'. These are defined (respectively) as:

$$\mathrm{GM}^{r,s}(x) = \mathrm{pol}_1(x) + \exp\{\mathrm{pol}_2(x)\} \tag{2.2}$$

where $\mathrm{pol}_1(x)$ and $\mathrm{pol}_2(x)$ are polynomials of order r and s respectively, and

$$\mathrm{LGM}^{r,s}(x) = \frac{\mathrm{GM}^{r,s}(x)}{1 + \mathrm{GM}^{r,s}(x)}. \tag{2.3}$$

These definitions enable a certain amount of flexibility to be achieved in the graduations by using either $\mathrm{GM}^{r,s}(x)$ or $\mathrm{LGM}^{r,s}(x)$ and varying the values of r and s. Forfar *et al.* (1988) investigate the choice of the 'best' model of this type in great detail. A major consideration is the trade-off between model fit (which can be improved by adding parameters) and parisomony (which requires fewer parameters).

It is natural to use an LGM formula for graduating q_x and a GM formula for graduating μ_x, as can be seen by considering the likelihood in each case.

To illustrate this, consider q_x.

$$P(A_x;q_x) = \binom{R_x}{A_x} q_x^{A_x}(1 - q_x)^{R_x - A_x}$$

$$= \binom{R_x}{A_x} \exp\{A_x \log\left(\frac{q_x}{1 - q_x}\right) + R_x \log(1 - q_x)\}. \tag{2.4}$$

This can be viewed as a member of an exponential family (which is the basis of generalised linear models).

$\mathrm{Log}(q_x/(1 - q_x))$ is known as the natural parameter, and it is sensible to apply the regression curve to this, rather than to q_x. One strong motivation for this choice can be appreciated by considering the range of values each function can take. Consider a regression curve, which is a function of age. This may be denoted, in vector notation, as $\boldsymbol{F}'\boldsymbol{\theta}$, where $\boldsymbol{F}$ is a column vector containing the regressor variables and $\boldsymbol{\theta}$ is a column vector containing the parameters. For example, the straight line $(\alpha + \beta x)$ has:

$$\boldsymbol{F} = \begin{bmatrix} 1 \\ x \end{bmatrix} \quad \text{and} \quad \boldsymbol{\theta} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$

It is clear that $\boldsymbol{F}'\boldsymbol{\theta}$ can take any value on the real line, depending on the values of the parameters α and β, and on the value of x. The probability of death q_x, however, must take a value between 0 and 1, and it would clearly lead to inconsistencies to model q_x by the regression function.

$(q_x/(1-q_x))$ can take any value on the positive half of the real line, and $\log(q_x/1-q_x)$ can take any value on the real line.

Thus, it appears sensible to model $\log(q_x/(1-q_x))$ by a regression function:

$$\log(q_x/(1-q_x)) = \boldsymbol{F}'\boldsymbol{\theta}. \tag{2.5}$$

The form of the regression function has to be determined and the parameters estimated.

Equation (2.5) can be rearranged as:

$$q_x = \frac{\exp\{\boldsymbol{F}'\boldsymbol{\theta}\}}{1+\exp\{\boldsymbol{F}'\boldsymbol{\theta}\}}. \tag{2.6}$$

Two things should be noted about this. First, the right hand side must lie between 0 and 1, which is compatible with q_x. Secondly, it is an LGM formula. Thus, it is sensible to model q_x by:

$$q_x = \mathrm{LGM}^{0,s}(x). \tag{2.7}$$

This form of curve is often found to provide a good model for q_x, and, in most cases, the first model fitted has $s=2$. Thus a straight line is fitted to $\log(q_x/(1-q_x))$. If this does not provide a satisfactory fit, then a better model is sought by increasing s or introducing a non-zero value of r.

In the dynamic regression method, the model $\mathrm{LGM}^{0,2}(x)$ is, again, the starting point, and greater flexibility is acquired in ways other than increasing the order of the polynomials.

For graduation of μ_x, it is $\log \mu_x$ which is modelled by the regression function:

$$\log R_x\mu_x = \boldsymbol{F}'\boldsymbol{\theta}. \tag{2.8}$$

Readers with a knowledge of generalised linear models will recognise these models for q_x and μ_x as the natural models to use in each case. Further investigation of the use of generalised linear models can be found in Renshaw (1991). It is possible to use a large number of other models within the framework of generalised linear models, and these are explored in detail by Renshaw (1991).

To summarise, the simplest parametric models which are considered are straight lines applied to the appropriate functions of the mortality parameters. These models sometimes provide a satisfactory fit. They are applied to the data from the CMI Committee (1976) (CMI Report No. 2) on Male Pensioners, 1967–70. In other cases a straight line fit is clearly unsatisfactory, and higher order polynomials have to be considered. The example used in this paper, when this is the case, is taken from Forfar *et al.* (1988), and consists of data on male assured lives. These data sets form the examples in Section 4.

It can be seen that graduation using parametric curves can be viewed as an application of generalised linear models. These are an extension of standard linear regression models to cases in which the data do not have a normal distribution. The parameter vector, $\boldsymbol{\theta}$, is estimated using maximum likelihood methods, with the likelihood being defined by the modelling distribution. The quantity $\boldsymbol{F}'\boldsymbol{\theta}$ is known as the linear predictor and, in the theory of generalised linear models, is related to the expected value of the observed variable via the link function. The computer package GLIM (Baker & Nelder, 1978) has been designed for this purpose, and provides a convenient framework for graduating data using this type of model.

The theory of generalised linear models in relation to graduation can be found in Renshaw (1991). The purpose of this paper is to extend the theory to dynamic generalised linear models, which are the subject of the next section.

3. Dynamic Models for Graduation

In this section the theory of generalised linear models for graduation, which was outlined in the previous sections, will be extended to allow for dynamic variation in the parameter vector. The standard regression line can be extended so that the parameters are not constant. Consider, first, a simple example, the straight line regression for normally distributed data:

$$y_i = \alpha + \beta x_i + e_i \qquad i = 1,2, \ldots, n. \tag{3.1}$$

Note that the data are indexed here by i, with the regressor variable being x, since this is the familiar form for regression. For graduation, the index variable will be the age, x.

The intercept α, and the gradient β, of the straight line are assumed to be constant for all i. Thus α and β do not depend on i.

Suppose now that the observations and the regressors are ordered (as is the case for graduation, when the data are naturally ordered with age). The parameters can be estimated by maximum likelihood estimation or, equivalently, by recursive maximum likelihood estimation. The recursive estimation method can be extended using the Kalman filter to allow the parameters to vary with i. There are two extreme cases which can easily be identified. The first, which has already been considered, is the 'static' regression method (which is the standard regression), in which the parameters are the same over the whole range of ages. The second is the case in which the parameters are estimated separately at each age. In other words, there are separate parameters for each i. This would produce the crude mortality rates. Between these two extremes are the cases when the parameters are similar at adjacent ages, but not identical. This allows, for example, a graduation which is locally linear, but not globally linear, to be fitted. When the fitted curve of this type is viewed over a small number of adjacent ages it appears to be straight. However, when the whole graduated curve is viewed, it is clear that the fitted line is not straight.

As is well known, some data are well graduated using a straight line, but in many cases a straight line fit is unsatisfactory. Two methods of overcoming this are either to use a higher order polynomial, or a non-linear function instead of a straight line, or to abandon a parametric curve and use a smoothing operation. The method in this paper is a combination of the two methods. The parametric curve is retained (usually as a straight line), and some smoothing is introduced. Thus, the parametric curve is estimated at each age, and a sophisticated smoothing procedure is used to relate adjacent ages.

The smoothing procedure uses the Kalman filter, which can be approached via Bayes theorem, and is usually applied when the data can be assumed to be normally distributed. The Kalman filter for normally distributed data is now summarised, and an indication of its extension to non-normally distributed data is given. The purpose of this paper is to illustrate the use of dynamic regression methods, rather than to re-derive them in great detail. The interested reader will find many derivations of the Kalman filter and its extensions in the papers listed among the references.

A fundamental and essential part of the present examination syllabus of the Institute of Actuaries is the derivation of the posterior distribution of the mean of a normal distribution, given an observed datum and the prior distribution. This forms part of the Kalman filter and the result is stated below.

Suppose: $\quad Y|\theta \sim \mathrm{N}(\theta,\sigma^2) \quad$ and $\quad \theta \sim \mathrm{N}(\mu,\sigma_0^2)$.

Then the posterior distribution of θ, given an observed value of Y, y is:

$$\theta|y \sim \mathrm{N}(\mu_*, \sigma_*^2) \tag{3.2}$$

where: $$\mu_* = \frac{\sigma_0^2 y + \sigma^2 \mu}{\sigma_0^2 + \sigma^2} \quad \text{and} \quad \sigma_*^2 = \frac{\sigma_0^2 \sigma^2}{\sigma_0^2 + \sigma^2}.$$

Of course, the usual situation is to have a sample of size n. The usual prior-posterior analysis has to be reformulated so that the data can be analysed recursively. Suppose $(i-1)$ observations have been analysed and the distribution of the mean, which is to be used as the prior distribution for the ith observation, is:

$$\theta \sim \mathrm{N}(a_i,P_i). \tag{3.3}$$

The conditional distribution of the ith observation is:

$$Y|\theta \sim \mathrm{N}(\theta,\sigma^2). \tag{3.4}$$

The posterior distribution of θ, given y_i, is:

$$\theta|y_i \sim \mathrm{N}(m_i, C_i) \tag{3.5}$$

where: $$m_i = \frac{P_i y_i + \sigma^2 a_i}{P_i + \sigma^2} \quad \text{and} \quad C_i = \frac{P_i \sigma^2}{P_i + \sigma^2}.$$

This posterior distribution can be used to formulate a prior distribution for the $(i+1)$th observation. There is considerable flexibility in this step, and it is the important ingredient in the Kalman filter. If the new prior distribution is the same as the previous posterior distribution, then the procedure is exactly as usual, and a recursive version of Bayes theorem for a sample of size n results.

i.e. $$a_{i+1} = m_i \qquad P_{i+1} = C_i. \tag{3.6}$$

This is equivalent to assuming that θ does not vary with i. Considerably more flexibility can be achieved by allowing θ to vary with i. Thus a suffix i must be attached to θ, and the recursion completed by defining the conditional distribution of θ_{i+1} given θ_i. This distribution is normal with mean $h\theta_i$. h is inserted to allow known changes in the mean to be modelled; often h will be unity. There is an increase in the variance to allow dynamic variation of the parameters from i to $(i+1)$. If there were no increase in the variance, recursive Bayes estimation would be obtained (as before).

Thus, the prior distribution for processing the $(i+1)$th datum is obtained from the posterior distribution from the ith datum as follows:

$$a_{i+1} = hm_i \qquad P_{i+1} = h^2 C_i / b^2. \tag{3.7}$$

h can be taken to be unity at present: its purpose will become clearer later.

b $(0 \leqslant b \leqslant 1)$ is known as the discount factor. If $b = 1$, there is no stochastic variation in the parameters and the static regression model is obtained. If $b = 0$, it can be seen that the variance of each prior distribution is infinite. Thus no information is passed from point to point, and the model is estimated separately at each point. A value of b between 0 and 1 gives a smoothed model.

The recursion is now complete and the next observation can be processed.

Usually the parameter is not a scalar, but a vector. In this case P_i is a variance-covariance matrix, h is a matrix and the discount factors are contained in a matrix.

In the applications in this paper the data are not normally distributed. In this case the Kalman filter has to be extended. Linear Bayes estimation is used, in a similar way as in Credibility theory, to obtain the recursive estimation equations which are given below. Further details of the derivation of these equations can be found in West *et al.* (1985) and Durbin (1990). The relevant recursions for the Poisson and binomial models are simply stated here, and no derivations are given.

For the data to be used in the graduation, the index is age x. Thus i is now replaced by x in the recursions. Recall that a parametric curve is fitted to the appropriate function of the mortality parameter at each age. At age x, this curve can be written as $\boldsymbol{F}'\boldsymbol{\theta}_x$, where $\boldsymbol{\theta}_x$ is the parameter vector.

The prior distribution of $\boldsymbol{\theta}_x$, given all the information up to age $(x-1)$, has mean $\boldsymbol{a}_x$ and covariance matrix P_x.

Let $$\mu_x = \boldsymbol{F}'\boldsymbol{a}_x \qquad \text{and} \qquad v_x = \boldsymbol{F}' P_x \boldsymbol{F}.$$

The posterior distribution of $\boldsymbol{\theta}_x$, given the data up to the age x, has mean $\boldsymbol{m}_x$ and covariance matrix C_x, where:

(1) for the Poisson distribution:

$$\boldsymbol{m}_x = \boldsymbol{a}_x + \frac{P_x\,\boldsymbol{F}_x}{v_x}\log\left(\frac{1 + v_x\,A_x}{1 + v_x\,R_x\,e^{u_x}}\right) \tag{3.8}$$

$$C_x = P_x - P_x\,F\,F'\,P_x\,\frac{A_x}{1 + v_x\,A_x} \tag{3.9}$$

(2) for the binomial distribution:

$$\boldsymbol{m}_x = \boldsymbol{a}_x + \frac{P_x\,\boldsymbol{F}_x}{v_x}\log\left(\frac{1 + e^{u_x} + v_x\,A_x}{1 + e^{u_x} + (R_x - A_x)\,v_x\,e^{u_x}}\right) \tag{3.10}$$

$$C_x = P_x - \frac{P_x\,F\,F'\,P_x}{v_x}\left(1 - \frac{1}{1 + e^{u_x} + v_x\,A_x} - \frac{1}{1 + e^{-u_x} + v_x\,(R_x - A_x)}\right). \tag{3.11}$$

For both models, the prior distribution of $\boldsymbol{\theta}_{x+1}$ for the next age has mean $\boldsymbol{a}_{x+1}$ and covariance matrix P_{x+1} where:

$$\boldsymbol{a}_{x+1} = H\,\boldsymbol{m}_x \qquad \text{and} \tag{3.12}$$

$$P_{x+1} = B\,H\,C_x\,H'\,B. \tag{3.13}$$

H is a matrix which can be used to model known changes in the parameters, and B is a diagonal matrix containing the discount factors. The diagonal elements of B are the inverses of the discount factors.

This completes the forward recursion. The data are processed recursively, and a final estimate of the parameter vector is obtained. For graduation an estimate of the parameter vector at each age (using *all* the available information, not just information up to each age) is required. Thus, it is necessary to apply another recursive procedure to go back through the data and obtain an estimate of the parameter vector at each age, using *all* the data and not just at lower ages. The forwards recursion gives an estimate of the mortality parameter at age x, using the data from ages 1 to x only. The backwards recursion gives the estimate using the data from all of the ages. This is the estimate which can be compared with the static regression estimates obtained by Renshaw (1991) and Forfar *et al.* (1988). The details of the second recursion are given below.

Suppose there are N ages for which data are available. Denote the mean and covariance matrix of the estimate of the parameter vector using all N observed ages by $\mathbf{m}_x^N$ and P_x^N. Thus the final recursion gives $\mathbf{m}_x^N$ and P_x^N for the final age only. From this the other estimates may be obtained (recursively) from:

$$\mathbf{m}_x^N = \mathbf{m}_x + C_x H' P_{x+1}^{-1} (\mathbf{m}_{x+1}^N - \mathbf{a}_{x+1}) \tag{3.14}$$

$$C_x^N = C_x + C_x H' P_{x+1}^{-1} (P_{x+1} - C_{x+1}^N) P_{x+1}^{-1} H C_x. \tag{3.15}$$

The process is initialised by using a vague prior distribution for the first observation. In order to ensure convergence to a sensible result, it was found that the initial parameter vector should be chosen close to the maximum likelihood estimate (i.e. the estimate obtained using the static regression method of Renshaw [1991] and Forfar *et al.* [1988] for a straight line fit). The vague prior distribution is obtained by setting the initial covariance matrix of the parameter vector to be diagonal, with large entries in the diagonal elements. This means that there is no information assumed known before the graduation begins; the choice of the initial parameter vector is merely to ensure convergence and does not imply any prior knowledge.

4. Examples

Two examples are now given. The first illustrates the dynamic estimation method in the case in which, in a conventional analysis, a straight line has been found to provide a satisfactory fit. In this case the discount factors are chosen close to 1, so that the model assumes that the line will be almost constant. This example illustrates the use of dynamic estimation when graduating the probability of death. The second example provides a case in which a straight line fit is unsatisfactory (when a static model is applied), and higher order terms have to be included. The dynamic straight line approach, in which the parameters are allowed to vary with age, provides a satisfactory fit.

The data for these examples are taken from the following sources:

(i) CMI Committee (1976) (C.M.I. Report No. 2). Graduation of the probability of death for Male Pensioners, 1967–70.

(ii) Forfar *et al.* (1988) Graduation of the force of mortality for Male Assured Lives for Duration 0.

The form of the dynamic linear predictor which has been fitted in each of these examples is:

$$g(\mu_x) = (\text{level})_x \tag{4.1}$$

where:

$$(\text{level})_x = (\text{level})_{x-1} + (\text{growth})_{x-1} \qquad (+\ \text{error}) \tag{4.2}$$

$$(\text{growth})_x = (\text{growth})_{x-1} \qquad (+\ \text{error}). \tag{4.3}$$

Thus the linear predictor consists of a single quantity, the 'level' at age x. The level at age x is related to the level at the previous age by adding the 'growth'.

This, it can be seen, is equivalent to fitting a straight line, in which the value increases by the gradient from age to age.

The growth is the same as the gradient of the straight line, and the intercept can be calculated from the level and growth at each age. The relationships are given in each example.

(i) *Example 1. Male Pensioners, 1967–70 (C.M.I. Report No. 2) (1976)*

The conventional analysis, using maximum likelihood estimation, can be found on page 76 of the C.M.I. Report. The model fitted is an LGM$^{(0,2)}$ and is given by the following formula:

$$\log\left(\frac{q_x}{1-q_x}\right) = -2\cdot9718602 + 4\cdot2142613\left(\frac{x-70}{50}\right). \tag{4.4}$$

The number of decimal places quoted appears rather optimistic in the light of the data and the estimation, and fewer are quoted in Tables 4.1–4.3.

In the dynamic estimation method, the straight line fitted is:

$$\log\left(\frac{q_x}{1-q_x}\right) = \alpha_x + \beta_x\left(\frac{x-70}{50}\right) \tag{4.5}$$

where the parameters, α_x and β_x, are allowed to vary with x. The amount of dynamic variation allowed is restricted by choosing both discount factors to be 0·995. If they were both taken as 1, the parameters would be constant. By choosing values very slightly less than 1, the lines are constrained to be almost straight, but some variation is allowed if the evidence from the data is very strong.

Equations (4.1)–(4.3) can be compared with equation (4.5) to give the following relationships between the parameters:

$$\alpha_x = (\text{level})_x + (70-x)\,(\text{growth})_x \tag{4.6}$$

$$\beta_x = 50\,(\text{growth})_x. \tag{4.7}$$

Table 4.1 shows the results of this graduation. In this table, A_x is actual deaths, and the quantity $(A_x - E_x)^2/E_x$ is required for the calculation of the χ^2 test statistic (with some regrouping where the expected number of deaths is low).

It can be seen that the parameters of the dynamic straight line are close to those of the static model throughout most of the ages considered. The only significant departures are in the early ages, and it should be noted that here the data are rather sparse.

The value of the χ^2 test statistic is 65·04, compared with 72·74 for the straight line fitted in the CMI Report. Thus, it can be seen that the dynamic graduation method gives a satisfactory fit.

Table 4.1. Dynamic Graduation of q_x for Male Pensioners, 1967–70

Age x	Level	Growth	Graduated q_x	Expected deaths E_x	$(A_x - Ex)^2/E_x$	α_x	β_x
50·5	−4·5858	0·1074	0·0101	6·279	0·827	−2·491	5·372
51·5	−4·5132	0·1074	0·0108			−2·527	5·368
52·5	−4·3991	0·1110	0·0121			−2·456	5·552
53·5	−4·3250	0·0925	0·0131			−2·799	4·623
54·5	−4·2434	0·0885	0·0142			−2·872	4·423
55·5	−4·1559	0·0883	0·0154			−2·875	4·417
56·5	−4·0839	0·0842	0·0166	8·730	1·594	−2·947	4·212
57·5	−4·0076	0·0829	0·0179			−2·972	4·144
58·5	−3·9346	0·0814	0·0192	6·061	1·426	−2·999	4·069
59·5	−3·8536	0·0819	0·0208	19·507	0·114	−2·993	4·097
60·5	−3·7705	0·0827	0·0225	46·620	0·041	−2·985	4·134
61·5	−3·6867	0·0833	0·0244	62·827	0·975	−2·979	4·163
62·5	−3·6038	0·0834	0·0265	81·619	0·235	−2·978	4·170
63·5	−3·5199	0·0837	0·0288	105·907	3·091	−2·976	4·184
64·5	−3·4346	0·0841	0·0312	1287·770	5·251	−2·972	4·207
65·5	−3·3495	0·0844	0·0339	3057·388	12·232	−2·970	4·218
66·5	−3·2658	0·0841	0·0368	3370·503	1·647	−2·971	4·205
67·5	−3·1822	0·0840	0·0398	3413·458	1·001	−2·972	4·198
68·5	−3·0986	0·0839	0·0432	3305·724	1·580	−2·973	4·194
69·5	−3·0147	0·0839	0·0468	3135·664	1·578	−2·973	4·196
70·5	−2·9306	0·0840	0·0507	2925·508	0·707	−2·973	4·199
71·5	−2·8464	0·0841	0·0549	2706·207	0·010	−2·972	4·203
72·5	−2·7622	0·0841	0·0594	2522·645	1·087	−2·972	4·205
73·5	−2·6778	0·0841	0·0643	2360·665	1·205	−2·972	4·207
74·5	−2·5934	0·0842	0·0696	2221·924	5·257	−2·972	4·210
75·5	−2·5088	0·0842	0·0752	2094·505	1·218	−2·972	4·212
76·5	−2·4243	0·0843	0·0813	1944·609	0·392	−2·972	4·214
77·5	−2·3398	0·0843	0·0879	1784·364	0·280	−2·972	4·215
78·5	−2·2554	0·0843	0·0949	1606·085	0·001	−2·972	4·216
79·5	−2·1709	0·0843	0·1024	1412·004	0·071	−2·972	4·217
80·5	−2·0865	0·0843	0·1104	1243·667	2·230	−2·972	4·217
81·5	−2·0021	0·0843	0·1190	1070·281	0·441	−2·972	4·217
82·5	−1·9176	0·0844	0·1281	892·577	0·571	−2·972	4·218
83·5	−1·8333	0·0844	0·1379	741·084	0·001	−2·972	4·218
84·5	−1·7489	0·0844	0·1482	593·871	0·199	−2·972	4·218
85·5	−1·6645	0·0844	0·1592	478·115	4·257	−2·972	4·218
86·5	−1·5802	0·0844	0·1708	369·712	0·161	−2·972	4·218
87·5	−1·4959	0·0843	0·1830	279·678	0·143	−2·972	4·217
88·5	−1·4116	0·0843	0·1960	209·206	2·270	−2·972	4·217
89·5	−1·3273	0·0843	0·2096	153·223	0·116	−2·972	4·217
90·5	−1·2430	0·0843	0·2239	117·109	3·047	−2·972	4·217
91·5	−1·1586	0·0843	0·2389	78·246	0·097	−2·972	4·217
92·5	−1·0742	0·0843	0·2546	55·884	0·468	−2·972	4·217
93·5	−0·9899	0·0843	0·2709	35·222	0·648	−2·972	4·217
94·5	−0·9055	0·0844	0·2879	22·746	0·024	−2·972	4·218
95·5	−0·8211	0·0844	0·3055	14·513	2·923	−2·972	4·218
96·5	−0·7367	0·0844	0·3237	7·608	1·711	−2·972	4·218
97·5	−0·6523	0·0844	0·3425	5·480	3·662	−2·972	4·218
98·5	−0·5679	0·0844	0·3617	5·123	0·246	−2·972	4·218
99·5	−0·4836	0·0844	0·3814			−2·972	4·218

(ii) *Example 2. Male Assured Lives, Permanent, Duration 0, Forfar* et al. (*1988*)

The second example uses the data for Example 3 of Forfar *et al.* (1988), which begins on page 101. Table 17.4 on page 116 graduates μ_x using the function GM (2,2). Clearly a straight line fit is unsatisfactory, and the discount factors in the dynamic estimation method have to be chosen to be less than 1.

The parameters are not expected to change radically from age to age, and so large discount factors are appropriate, although less than in the first example. Values of 0·95 were chosen after some experimentation. It is clear that the fit can be improved by reducing the discount factors, but the smoothness of the fitted line is forsaken. This can be seen to be similar to the assessment of parsimony and

Table 4.2. Dynamic Graduation of μ_x for Male Assured Lives, Duration 0

Age x	Level	Growth	Graduated q_x	Expected deaths E_x	$(A_x - Ex)^2/E_x$	α_x	β_x
10	−10·014	0·008	0·00004			−9·5402	0·3947
11	−9·845	−0·002	0·00005			−9·9665	−0·1028
12	−9·564	0·003	0·00007			−9·3622	0·1742
13	−9·134	0·029	0·00011	20·4242 (ages 10–17)	0·2877 (ages 10–17)	−7·4715	1·4580
14	−8·482	0·079	0·00021			−4·0603	3·9477
15	−7·587	0·145	0·00051			0·4135	7·2730
16	−7·468	0·076	0·00057			−3·3833	3·7821
17	−7·396	0·049	0·00061			−4·7998	2·4491
18	−7·360	0·028	0·00064	19·5401	2·1356	−5·9135	1·3910
19	−7·395	−0·005	0·00061	23·2344	1·9701	−7·6746	−0·2742
20	−7·452	−0·023	0·00058	25·1314	0·1389	−8·5874	−1·1356
21	−7·498	−0·026	0·00055	27·2175	0·5257	−8·7563	−1·2840
22	−7·542	−0·026	0·00053	29·5709	0·1995	−8·7943	−1·3046
23	−7·579	−0·024	0·00051	30·6860	1·2992	−8·7272	−1·2209
24	−7·620	−0·024	0·00049	30·7973	1·6845	−8·7076	−1·1822
25	−7·664	−0·023	0·00047	30·1245	0·1498	−8·7139	−1·1669
26	−7·693	−0·020	0·00046	29·7375	0·0183	−8·5902	−1·0192
27	−7·713	−0·016	0·00045	28·8037	0·6113	−8·4167	−0·8186
28	−7·731	−0·013	0·00044	28·4203	4·5891	−8·2757	−0·6486
29	−7·722	−0·007	0·00044	28·5310	0·0756	−7·9910	−0·3279
30	−7·710	−0·001	0·00045	29·9036	0·8041	−7·7473	−0·0472
31	−7·684	0·005	0·00046	31·9363	1·9722	−7·4807	0·2604
32	−7·642	0·012	0·00048	34·1706	1·1143	−7·1830	0·6036
33	−7·588	0·019	0·00051	36·2895	0·1444	−6·8854	0·9495
34	−7·529	0·025	0·00054	36·4670	0·1669	−6·6175	1·2658
35	−7·464	0·031	0·00057	36·5833	0·0685	−6·3713	1·5611
36	−7·394	0·037	0·00061	36·0501	0·2414	−6·1464	1·8352
37	−7·324	0·041	0·00066	35·5435	0·1698	−5·9559	2·0732
38	−7·253	0·046	0·00071	34·3916	2·0476	−5·7901	2·2851
39	−7·170	0·050	0·00077	34·3780	0·1645	−5·6089	2·5171
40	−7·081	0·055	0·00084	35·5744	0·0051	−5·4363	2·7411
41	−6·990	0·059	0·00092	36·2286	1·0708	−5·2815	2·9457
42	−6·891	0·063	0·00102	37·3447	3·6376	−5·1241	3·1553
43	−6·800	0·066	0·00111	39·2448	2·6744	−5·0118	3·3108

Table 4.2 (*cont.*)

Age x	Level	Growth	Graduated q_x	Expected deaths E_x	$(A_x - Ex)^2/E_x$	α_x	β_x
44	–6·698	0·070	0·00123	40·8759	0·6423	–4·8864	3·4843
45	–6·599	0·073	0·00136	43·4433	1·2753	–4·7857	3·6276
46	–6·494	0·076	0·00151	43·3954	0·0449	–4·6815	3·7767
47	–6·387	0·078	0·00168	44·6929	0·6302	–4·5867	3·9143
48	–6·283	0·081	0·00186	46·2458	0·3048	–4·5119	4·0257
49	–6·181	0·082	0·00206	49·4725	1·4699	–4·4512	4·1179
50	–6·083	0·084	0·00228	54·3231	3·9654	–4·4095	4·1836
51	–5·993	0·084	0·00249	52·4568	0·1233	–4·3903	4·2166
52	–5·903	0·085	0·00272	47·9765	3·5099	–4·3756	4·2433
53	–5·806	0·086	0·00300	48·2639	0·0112	–4·3475	4·2913
54	–5·710	0·087	0·00330	49·0290	0·1800	–4·3230	4·3337
55	–5·614	0·087	0·00363	53·3410	0·0516	–4·3043	4·3668
56	–5·520	0·088	0·00399	45·4429	7·5780	–4·2894	4·3935
57	–5·434	0·088	0·00435	33·3260	2·0801	–4·2926	4·3886
58	–5·343	0·088	0·00476	27·4524	0·7221	–4·2878	4·3983
59	–5·250	0·088	0·00522	26·2625	4·3901	–4·2785	4·4163
60	–5·163	0·088	0·00569	27·1073	2·9173	–4·2805	4·4129
61	–5·082	0·088	0·00617	20·2606	0·5247	–4·2908	4·3929
62	–4·998	0·088	0·00670	14·5247	0·8554	–4·2974	4·3808
63	–4·913	0·088	0·00730	12·6592	0·0343	–4·2997	4·3775
64	–4·826	0·088	0·00795	13·8925	0·2578	–4·3012	4·3764
65	–4·738	0·088	0·00868	24·9461	0·6242	–4·2999	4·3814
66	–4·645	0·088	0·00951	20·5235	0·3103	–4·2936	4·3979
67	–4·550	0·088	0·01046	12·8168	0·0521	–4·2846	4·4208
68	–4·452	0·089	0·01152	9·9691	1·5802	–4·2746	4·4461
69	–4·349	0·090	0·01276	9·3016	0·7828	–4·2592	4·4845
70	–4·248	0·090	0·01410	8·6434	1·3035	–4·2477	4·5137
71	–4·150	0·091	0·01552	7·3083	0·0130	–4·2408	4·5312
72	–4·051	0·091	0·01710	6·5070	6·4792	–4·2334	4·5502
73	–3·962	0·091	0·01866	6·0565	1·5425	–4·2350	4·5452
74	–3·868	0·091	0·02048	5·2418	0·2942	–4·2320	4·5527
75	–3·770	0·091	0·02253	8·1698	0·9804	–4·2269	4·5661
76	–3·677	0·091	0·02466			–4·2256	4·5683
77	–3·585	0·091	0·02699			–4·2247	4·5694
78	–3·491	0·091	0·02956			–4·2230	4·5729
79	–3·411	0·091	0·03195			–4·2301	4·5494
80	–3·330	0·091	0·03455			–4·2361	4·5290
81	–3·248	0·090	0·03741			–4·2407	4·5132
82	–3·165	0·090	0·04052			–4·2445	4·4995
83	–3·081	0·090	0·04389	5·7246	0·2841	–4·2479	4·4875
84	–2·998	0·090	0·04753			–4·2510	4·4758
85	–2·915	0·089	0·05141			–4·2543	4·4637
86	–2·833	0·089	0·05558			–4·2574	4·4521
87	–2·751	0·089	0·06004			–4·2605	4·4405
88	–2·667	0·089	0·06493			–4·2628	4·4319
100	–1·734	0·086	0·15007			–4·3012	4·2786

Table 4.3. Comparison of Dynamic and Static Graduation of μ_x for Male Assured Lives, Duration 0

		Dynamic Model		Static Model	
Age	A_x	E_x	$A_x - E_x$	E_x	$A_x - E_x$
10	0	0·01	−0·01	0·25	−0·25
11	0	0·01	−0·01	0·25	−0·25
12	0	0·02	−0·02	0·31	−0·31
13	0	0·06	−0·06	0·59	−0·59
14	0	0·14	−0·14	0·63	−0·63
15	1	0·63	0·37	1·10	−0·10
16	3	3·73	−0·73	5·44	−2·44
17	14	15·08	−1·08	19·18	−5·18
18	26	19·54	6·46	22·48	3·52
19	30	23·23	6·77	25·94	4·06
20	27	25·13	1·87	27·83	−0·83
21	31	27·22	3·78	29·60	1·40
22	32	29·57	2·43	31·56	0·44
23	37	30·69	6·31	32·01	4·99
24	38	30·80	7·20	31·61	6·39
25	28	30·12	−2·12	30·68	−2·68
26	29	29·74	−0·74	29·83	−0·83
27	33	28·80	4·20	28·40	4·60
28	17	28·42	−11·42	27·79	−10·79
29	30	28·53	1·47	27·24	2·76
30	25	29·90	−4·90	28·15	−3·15
31	24	31·94	−7·94	29·65	−5·65
32	28	34·17	−6·17	31·23	−3·23
33	34	36·29	−2·29	32·73	1·27
34	34	36·47	−2·47	32·71	1·29
35	35	36·58	−1·58	32·85	2·15
36	39	36·05	2·95	32·60	6·40
37	38	35·54	2·46	32·64	5·36
38	26	34·39	−8·39	32·26	−6·26
39	32	34·38	−2·38	32·75	−0·75
40	36	35·57	0·43	34·36	1·64
41	30	36·23	−6·23	35·51	−5·51
42	49	37·34	11·66	36·91	12·09
43	29	39·24	−10·24	39·45	−10·45
44	46	40·88	5·12	41·40	4·60
45	36	43·44	−7·44	44·43	−8·43
46	42	43·40	−1·40	44·50	−2·50
47	50	44·69	5·31	45·82	4·18
48	50	46·25	3·75	47·49	2·51
49	58	49·47	8·53	50·91	7·09
50	69	54·32	14·68	56·20	12·80
51	55	52·46	2·54	54·88	0·12
52	35	47·98	−12·98	50·73	−15·73
53	49	48·26	0·74	51·14	−2·14
54	52	49·03	2·97	51·98	0·02
55	55	53·34	1·66	56·56	−1·56
56	64	45·44	18·56	48·17	15·83
57	25	33·33	−8·33	35·57	−10·57

Table 4.3 (*cont.*)

		Dynamic Model		Static Model	
Age	A_x	E_x	A_x-E_x	E_x	A_x-E_x
58	23	27·45	−4·45	29·34	−6·34
59	37	26·26	10·74	27·99	9·01
60	36	27·11	8·89	28·95	7·05
61	17	20·26	−3·26	21·78	−4·78
62	11	14·52	−3·52	15·67	−4·67
63	12	12·66	−0·66	13·66	−1·66
64	12	13·89	−1·89	14·97	−2·97
65	21	24·95	−3·95	26·76	−5·76
66	18	20·52	−2·52	21·81	−3·81
67	12	12·82	−0·82	13·44	−1·44
68	6	9·97	−3·97	10·29	−4·29
69	12	9·30	2·70	9·39	2·61
70	12	8·64	3·36	8·54	3·46
71	7	7·31	−0·31	7·09	−0·09
72	13	6·51	6·49	6·19	6·81
73	3	6·06	−3·06	5·70	−2·70
74	4	5·24	−1·24	4·85	−0·85
75	8	5·47	2·53	4·97	3·03
76	3	2·70	0·30	2·41	0·59
77	1	1·37	−0·37	1·20	−0·20
78	6	1·40	4·60	1·21	4·79
79	0	0·75	−0·75	0·65	−0·65
80	0	0·78	−0·78	0·67	−0·67
81	0	0·39	−0·39	0·33	−0·33
82	0	0·26	−0·26	0·22	−0·22
83	0	0·13	−0·13	0·11	−0·11
84	0	0·05	−0·05	0·04	−0·04
85	0	0·10	−0·10	0·08	−0·08
86	0	0·06	−0·06	0·05	−0·05
87	0	0·18	−0·18	0·15	−0·15
88	0	0·10	−0·10	0·08	−0·08
100	0	0·15	−0·15	0·12	−0·12

the use of an information criterion, as is discussed further in the final section. The choice of discount factors produces a fit approximately as good as that obtained by Forfar *et al.* (1988). Table 4.2 shows the results of the dynamic graduation.

Table 4.2 can be compared with the table on pages 116–17 of Forfar *et al.* (1988). The value of the χ^2 test statistic (after regrouping to ensure all expected numbers of deaths are greater than 5) is 73·8, compared with 71·14 that they obtained. Thus the fits are almost equally as good on this basis. It can be seen by observation that the graduated values are quite similar to those in Forfar *et al.* (1988), and thus the non-parametric tests they used would also yield similar results. The smoothness and the tightness of the fit can be adjusted easily for the dynamic regression procedure by altering the discount factors.

The gradient of the dynamic straight line is roughly constant after about age 50, and it is negative between ages 20 and 30. For the early ages, the parameters are rather erratic due to the sparsity of the data.

Table 4.3 compares the dynamic graduation results with the static version in Forfar *et al.* (1988).

It is clear that the dynamic graduation method, with the discount factors relatively large (but nevertheless not equal to 1) gives a satisfactory fit. There is thus quite a lot of smoothing in this graduation.

Only the χ^2 test statistic has been quoted in these examples. Further details of the graduation tests may be found in the original papers. It is possible to obtain better fits by altering the discount factors. In this case, the size of the χ^2 test static can be controlled, to a certain extent, by the graduator. A further discussion of this is given in the next section.

5. Comments and Conclusions

In practical applications, it may be necessary to extend a graduated mortality table to ages outside the range of the data. This is one advantage of using a mathematical formula, although such a formula does not necessarily produce 'sensible' projections. One way of dealing with the problem, using the dynamic regression method, is simply to project a straight line using the parameters derived for the lowest or highest age, as appropriate. This might well produce reasonable results for the higher ages, but one cannot be sure that a sensible projection will be found. Previous smoothing procedures applied to graduation include the Whittaker–Henderson method, described in Benjamin & Pollard (1980) and that derived by Copas & Haberman (1983). The dynamic regression method shares with these methods the element of sequential smoothing, but it also contains the basic foundation of parametric curve fitting. It can be regarded as a combination of these two procedures. The amount of smoothing is governed by the discount factors, and is at present decided entirely by the modeller. However, the data obviously influence the choice, though not in an automatic way such as is implicit in the use of Akaike's information criterion, AIC (for example). The author of this paper is investigating the possibility that the information criteria can be extended in order to be of use in this situation. Other authors have suggested using a criterion similar to AIC, but with an arbitrary choice of weighting for the goodness of fit and the number of parameters. The choice of the weighting by the modeller replaces the choice of the amount of smoothing, and the procedure is again not automatically determined by the data. The Whittaker–Henderson method makes explicit the trade-off between smoothness and goodness of fit. The Whittaker–Henderson formula penalises the least-squares criterion by adding a term which depends on the smoothness of the curve. Using the notation of Section 11.71 of Benjamin & Pollard, the function to be minimised becomes:

$$Q = \sum_{j=0}^{n} w_j \, (\hat{q}_{x+j} - \overset{0}{q}_{x+j})^2 + K \sum_{j=0}^{n-3} (\Delta^3 \hat{q}_{x+j})^2 \tag{4.8}$$

where $\hat{q}$ are the fitted values, and $\overset{0}{q}$ are the crude values.

The first term in equation (4.8) is the least-squares quantity, and the second term, which involves the third differences of the fitted values, measures the smoothness of the fitted curve. Thus, the first term can be decreased by choosing fitted values close to the crude values, resulting in a less smooth fit, and the second term can be decreased by departing from the crude values in order to make the fit as smooth as possible. Minimising Q results in a compromise between the two terms, and hence between fit and smoothness. The relative importance of the two terms in equation (4.8) is determined by the choice of K. This has to be chosen subjectively by the graduator, and the similarity between the choice of K and the choice of the discount factors is clear.

Copas & Haberman (1983) apply non-parametric methods using a kernel function. The simplest example of a kernel function given by Copas & Haberman (1983) is:

$$\psi_s(u) = \begin{cases} 0 & u \leqslant -h \\ u + h & -h \leqslant u \leqslant 0 \\ h - u & 0 \leqslant u \leqslant h \\ 0 & u \geqslant h \end{cases}$$

Functions of this type can be used for graduation, and the degree of smoothing is controlled by h. When h is small the closeness of the fit takes precedence, and as h becomes larger, the smoothness increases. Again, the value of h has to be chosen by the graduator. In the words of Copas & Haberman, "the choice of h . . . is basically a subjective one, involving a compromise between reflecting important features of the data and yet not over-reacting to spurious chance fluctuations".

In the end, it seems sufficient to choose the discount factors by observation, with values in the region (0·9, 1·0) appearing sensible. It would be more elegant to have a theoretical version of AIC available for this purpose.

This paper has combined the parametric curve-fitting method of graduation with a method based on smoothing. As was noted by Copas & Haberman (1983), "the degree to which the estimated curve responds to features of the data can be controlled in a continuous fashion. By contrast, the smoothness of parametric methods can only be regulated in discrete steps".

REFERENCES

BAKER, R. J. & NELDER, J. A. (1978). *GLIM Release 3*. Nag, Oxford.

BENJAMIN, B. & POLLARD, J. H. (1980). *The Analysis of Mortality and Other Actuarial Statistics*. Heinemann, London.

C.M.I. COMMITTEE (1976). The Graduation of Pensioners' and of Annuitants' Mortality Experience, 1967–70, *CMIR*, **2**, 57.

COPAS, J. B. & HABERMAN, S. (1983). Non-Parametric Graduation using Kernel Methods. *J.I.A.* **110**, 135.

DURBIN, J. (1990). Extensions of Kalman Modelling to non-Gaussian Observations. *Quaderni di Stat. e Math. Applic. alle Scienze Econ. Sociali*, **XII**, 1, 3.

FORFAR, D. O., MCCUTCHEON, J. J. & WILKIE, A. D. (1988). On Graduation by Mathematical Formula. *J.I.A.* **115**, 1 and *T.F.A.* **41**, 97.

MCCULLAGH, P. & NELDER, J. R. (1989). *Generalised Linear Models*, 2nd ed. Chapman and Hall.

RENSHAW, A. E. (1991) Actuarial Graduation Practice and Generalised Linear and Non-Linear Models. *J.I.A.* **118**, 295.

WEST, M., HARRISON, P. J. & MIGON, H. S. (1985). Dynamic Generalised Linear Models and Bayesian Forecasting, with discussion. *J. American Statist. Ass.*, **80**, 73.