JIA 118 (1991) 453-481

J.I.A. **118**, III, 453–481

ON MODELLING SELECT MORTALITY

BY I. D. CURRIE, B.Sc., Ph.D., F.S.S.
AND H. R. WATERS, M.A., D.Phil., F.I.A.

ABSTRACT

In this paper we present an approach to the graduation of select mortality data and we illustrate this approach by graduating the CMIB's data for Female Permanent Assurances 1979–1982. The difference between our approach and that of the CMIB is that we graduate simultaneously by attained age and duration since selection.

KEYWORDS

Graduation; Mortality

1. INTRODUCTION

The Continuous Mortality Investigation Bureau (CMIB) published in *CMIR* **9** (1988) graduations of many sets of data relating to the years 1979–1982. The purpose of this paper is to present a different approach to the graduation of select mortality data, illustrating the method by graduating just one of the CMIB's sets of data, Permanent Assurances, Females. Our approach to the graduation of select mortality data is to graduate *simultaneously* by (attained) age and duration since selection, rather than to graduate by age *separately* for each duration. There are two reasons why we consider this exercise to be interesting and useful:

(a) it makes more efficient use of the available data than the traditional method of graduation, since it allows us to infer information about ages and durations where we have little data, from other ages and durations where we have more data, and
(b) by modelling the effect of duration since selection, we ought to gain more insight into the nature of selection.

It should be stressed that it was not our purpose to produce a 'law' of (select) mortality in the spirit of, say, Gompertz' law of mortality, or of Tenenbein & Vanderhoof (1980). It has been our purpose merely to smooth the data, though by fitting a (two-dimensional) surface rather than a series of (one-dimensional) curves. Some recent work by Panjer & Russo (1990) has a similar objective to ours, though their data and their approach are somewhat different to ours. The comments by Lyons (1990) on the work of the CMIB endorse the view that our objective is a reasonable, and perhaps even an obvious, one.

In Section 2 we give a description of the data available to us, and in Section 3 we give a brief summary of the CMIB's graduations of these data. Our own method of graduation is described in Section 4 and the results are given in Section 5. (The graduation in Section 5 is the last of a long list of graduations which were tried and (apart from the final one) rejected. This process took a not inconsiderable amount of time, during which our already high regard for the work of the CMIB increased considerably!) The graduation in Section 5 required some 'hand smoothing' to make it acceptable when extrapolated beyond the ranges of the data, and these adjustments are described in Section 6. In Section 7 we make some comments about the effect on the force of mortality of duration since selection and in Section 8 we make some comparisons between our graduation and those of the CMIB. We make a few concluding remarks in Section 9.

We are grateful to the CMIB for making data available to us, and also to individual members of the CMIB, in particular our colleagues John McCutcheon and David Wilkie, for help and advice throughout the course of this project. The research for the paper was partly supported by The Royal Society and The Carnegie Trust for the Universities of Scotland and the first author gratefully acknowledges this financial support which allowed a preliminary version of the paper to be presented to the 24th ARCH Conference in 1989; the paper greatly benefited from some comments made by Gary Venter at the conference. We are grateful to an anonymous referee for some helpful comments on an earlier draft of this paper.

2. THE DATA

The data used for our graduation are the CMIB's data for Permanent Assurances, Females, in the years 1979–1982. These include lives who have not been medically examined as well as those who have, but only include lives accepted at normal premium rates.

The data supplied to us by the CMIB consisted of values of:

$\theta(d,x)$ for $x = 10, 11, \ldots, 108$ and $d = 0, 1, \ldots, 4$, where $\theta(d,x)$ is the number of deaths at (attained) age x nearest birthday and curtate duration d,

$E(d,x)$ the central exposure corresponding to $\theta(d,x)$,

$\theta(5+,x) \equiv \sum_{d=5}^{\infty} \theta(d,x)$ for $x = 10, 11, \ldots, 108$,

$E(5+,x) \equiv \sum_{d=5}^{\infty} E(d,x)$ for $x = 10, 11, \ldots, 108$,

m_1 and m_2 first and second moments about zero, respectively, of the number of policies per individual life for each age x, $x = 20, 21, \ldots, 110$, separately for durations 0, 1, 2, 3, 4, 5+.

Note that for each attained age x, the variance ratio, denoted r_x, is calculated as:

$$r_x = m_2/m_1.$$

The role of the variance ratio is explained fully by Forfar, McCutcheon & Wilkie (1988). After examining the data on duplicate policies, which are the results of an investigation based on deaths rather than on policies, we came to the conclusion that there was little evidence of duplicate policies for single durations, though there was evidence of duplicate policies in the duration 5+ data. See the comment in the final paragraph of Section 6.1 of *CMIR* **9** (1988).

Some statistics relating to these data are given in *CMIR* **9** (1988, Section 6.1) from which it is clear that for durations less than 5 years there are relatively few data, in terms of the number of deaths, outside the age range 45–60 (see comment (a) in our Section 1). Table 1 shows the deaths and exposures for each duration summed, for convenience, into 5-year age groups from 20 to 90. Table 2 shows, for selected ages, the average number of policies for an individual life, m_1, and the variance ratio, r_x, for the duration 5+ data. A noticeable feature of Table 2 is the very high value of r_{85}. (This was in fact the highest value of r_x in the data.) We used the values of r_x, in particular the value of r_{85}, without making any attempt at smoothing, for two reasons. First, the r_x values were estimated by the CMIB using a very large sample, and second, the r_x values do not affect the point estimates of the force of mortality, but only the weights to be attached to these estimates.

It was inconvenient for our purposes to have, for each age x, the data for durations 5 and above summed into single values for $\theta(5+,x)$ and $E(5+,x)$. We would have preferred to have values of $\theta(d,x)$ and $E(d,x)$ for $d=5,6,\ldots$, but these were not available to us, and we were unable to estimate them. What we were able to estimate was a function $D(x)$, which represents for attained age x the average duration of policy for those policies whose duration exceeds 5 years. The details of our method for estimating $D(x)$ are given in the Appendix and values of $D(x)$ are shown in Figure 1. The estimation of $D(x)$ involves a number of assumptions and approximations, but the results, as shown in Figure 1, are not, in our opinion, unreasonable. For example, apart from a marked drop just after age 60, the average duration increases with age; at age 50 the average duration (for policies which have been in force at least 5 years) is about 10 years. The drop in the average duration between ages 59 and 63 could be caused by a large number of females effecting, at a young age, policies which mature at or around their 60th birthday.

3. THE CMIB'S GRADUATIONS

The CMIB's general approach to the graduation of select mortality data is to graduate separately each of the curtate durations since selection 0, 1, 2, 3, 4 and 5+ years, with the possibility that some of the durations could be combined to

Table 1. *The data summed into 5-year age groups*

	Curtate Duration					
	0		1		2	
Age	Exposures	Deaths	Exposures	Deaths	Exposures	Deaths
20–24	120276	18	105301	30	82547	17
25–29	116589	29	109487	28	96635	31
30–34	120652	28	114904	34	102733	31
35–39	96696	39	93893	56	86006	45
40–44	73863	37	71023	53	65275	46
45–49	59871	49	57706	75	53902	79
50–54	43419	88	44167	99	42961	102
55–59	21001	47	22126	90	23070	108
60–64	9640	27	8938	46	8764	57
65–69	4190	20	3877	39	3583	34
70–74	1812	22	1495	27	1269	11
75–79	480	1	558	13	499	5
80–84	71	1	84	1	91	1
85–90	6	0	9	0	11	1
Total	668563	406	633565	591	567345	568

	3		4		5+	
Age	Exposures	Deaths	Exposures	Deaths	Exposures	Deaths
20–24	64394	13	44759	10	54323	16
25–29	84640	28	71921	24	187983	68
30–34	89745	28	76367	25	295413	128
35–39	76613	33	65334	39	290985	189
40–44	58937	50	50937	41	240867	266
45–49	48857	73	42825	73	228286	404
50–54	40297	95	35948	92	223775	756
55–59	23258	93	22653	103	194852	951
60–64	8860	59	9137	47	93376	653
65–69	3334	38	3057	25	27005	277
70–74	1150	6	1166	19	12111	258
75–79	448	6	390	10	6135	205
80–84	90	1	96	4	3250	226
85–90	15	0	15	1	1718	210
Total	500636	523	424602	513	1860077	4607

give a graduation of, say, the data for durations 2+ years (≡2, 3, 4 and 5+ years).

The methodology used by the CMIB to graduate the 1979–1982 data sets is explained very fully in Forfar, McCutcheon & Wilkie (1988) and the results of the graduations are given in *CMIR* **9** (1988). Briefly, the method used was to graduate the force of mortality by mathematical formula, with the parameters fitted by maximum likelihood. The mathematical formula was chosen from two families of functions of which, for the Permanent Assurances, Females, graduations, the most generally accepted type of function for μ_x was of the form:

Table 2. *Estimates of the average number of policies per life and of the variance inflation factor in duration 5+ data*

Age	Average number, m_1, of policies per life	Variance inflation factor r_x
20	1·000	1·00
25	1·118	1·21
30	1·083	1·15
35	1·139	1·39
40	1·032	1·06
45	1·088	1·16
50	1·085	1·16
55	1·092	1·21
60	1·094	1·23
65	1·074	1·14
70	1·057	1·16
75	1·045	1·09
80	1·087	1·16
85	1·333	3·25
90	1·000	1·00

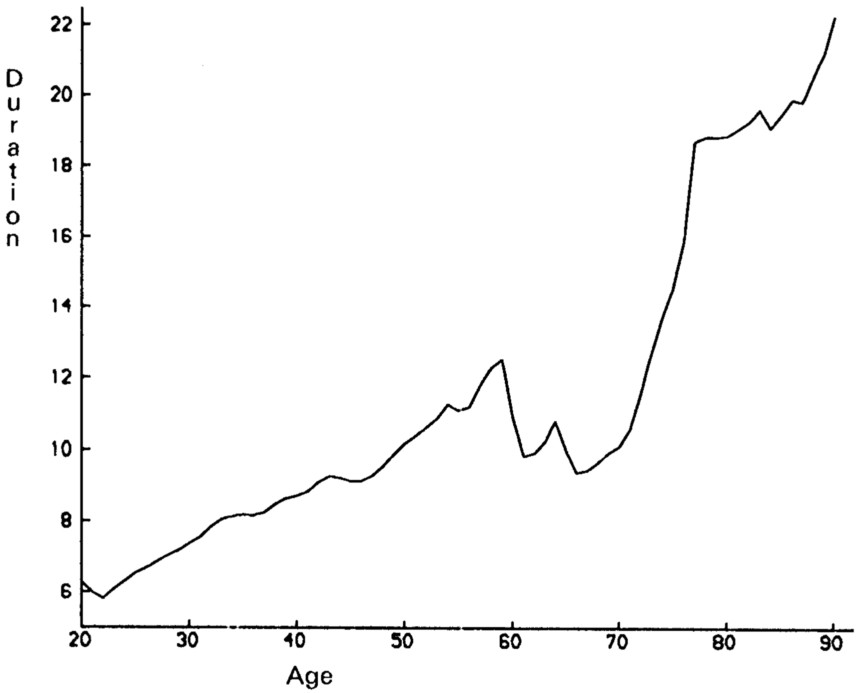

Figure 1. The function $D(x)$.

$$a_0+a_1x+\exp\{b_0+b_1x\}$$

where x denotes attained age and a_0, a_1, b_0 and b_1 are parameters to be fitted, i.e. a GM(2,2) function in the notation of Forfar, McCutcheon & Wilkie (1988). This function has 4 parameters so that graduations of all 6 durations involve a total of 24 parameters, although a graduation of just the 0, 1 and 2+ durations would involve only 12 parameters.

One problem with graduating the durations separately is that apparent anomalies can occur. Examples can be found in Table 6.2 of *CMIR* **9** (1988) where we see that the values of q_x for ages 20, 30 and 70 and over are higher for the duration 1 graduation than for the durations 2–4 graduation, in both cases taking the GM(2,2) graduation. It should be noted that there are very little data for these combinations of age and duration (see *CMIR* **9** (1988, Section 6.1) and comment (a) in our Section 1).

Following discussions at the Faculty of Actuaries and the Institute of Actuaries, the CMIB has produced a final graduation of the Permanent Assurances, Females, 1979–1982 select mortality data. Their final graduation, to be known as the AF80 table, has a 2-year select period and is based on the GM(2,2) graduations of the durations 0, 1 and 2+ data given in *CMIR* **9** (1988), with some 'hand smoothing' at high and low ages. The AF80 table was published in *CMIR* **10** (1990) and we are grateful to the CMIB for permission to see and use their results before their publication.

4. THE STATISTICAL MODELS

The key element in our attempt to graduate select mortality data simultaneously for age and duration is our treatment of the 5+ data. As mentioned briefly in Section 2 we estimated, for each attained age x, the average duration of policy for those policies whose duration exceeded 5 years. We now assume that we can consider all the deaths and all the exposure for attained age x and durations in excess of 5 years to be concentrated at this average duration, $D(x)$. It is helpful to think of a point estimate of mortality occurring at age nearest birthday x and exact duration $D(x)$ and equal to $\theta(5+,x)/E(5+,x)$. Our method smooths these 'new data points' along with the original data for durations 0, 1, 2, 3 and 4.

We shall denote by $\mu_{x,d}$ the force of mortality at exact attained age x and exact duration since selection d, both measured in years. We wanted our model for $\mu_{x,d}$:

(i) to be monotonic increasing in age for fixed duration, except possibly at the younger ages;
(ii) to be monotonic increasing in duration for fixed age, the duration effect tending to a finite limit as $d\to\infty$;
(iii) to have a simple interaction term (if required); and
(iv) to allow the log link in a generalised linear model.

We considered models of the following form:

$$\mu_{x,d} = \exp\{f_1^*(x) + f_2^*(d) + f_3^*(x,d)\} \tag{4.1}$$

where the functions f_1^*, f_2^* and f_3^* satisfied the following conditions:

(a) The age term, $\exp\{f_1^*(x)\}$, was of the form GM $(0, s)$, i.e. $f_1^*(x) = p(x)$ where $p(x)$ is a polynomial of degree $(s-1)$ in x. For technical reasons, the effect of age was actually fitted using a transformed value z of x, where:

$$p(x) = \sum_{j=0}^{s-1} a_j z^j, \qquad \text{and } z = (x - 70)/50.$$

This is the same transformation of age as used by the CMIB.

(b) The duration term was of the form $\exp\{f_2^*(d)\}$, where $f_2^*(d) = q(d)$ was a polynomial of degree t in $1/(d+k)$ and k was further parameter introduced to improve the behaviour of $f_2^*(d)$ for small d.

(c) The interaction term, $f_3^*(x,d)$, was of a simple form.

The function (4.1) is used to describe the underlying systematic relationship between the force of mortality, $\mu_{x,d}$, and age, x, and duration, d. We will make some comments on this choice of function when we have described the random component of our model for the observed number of deaths. The observed variation about $\mu_{x,d}$ is accounted for by an extension of a Poisson model which we describe next.

Let θ denote the number of deaths between exact ages x and $(x+1)$, and at curtate duration d, and let E denote the corresponding central exposed to risk. Then, ignoring for the moment the problem of duplicates, we assume that the force of mortality is constant over the rectangle $[x, x+1) \times [d, d+1)$ and equal to μ, say. We can suppose that θ has a Poisson distribution with mean $E\mu$ and write $\theta \sim \mathrm{P}(E\mu)$. We can summarise our approach in the following way:

$$\theta \sim \mathrm{P}(E\mu), \text{ where } \mu \ (= \mu_{x+\frac{1}{2}, d+\frac{1}{2}}) \text{ is given by (4.1)} \tag{4.2}$$

It is worth noting that the approach of the CMIB can be thought of as fitting separate functions $\mu = \mu_x$ of the form GM (r,s) for each duration for some values, possibly different for each duration, of r and s.

We suppose for the moment that k in (b) above is known. Then, still ignoring the problem of duplicates, the model defined in (4.1) and (4.2) belongs to the class of models known as generalised linear models (see Nelder & Wedderburn (1972) and McCullagh & Nelder (1989); see also Currie (1990) and Renshaw (1991)). These models are extremely simple to fit using packages such as GLIM and GENSTAT; indeed, the computations for this project were all conducted using GENSTAT 5 (Genstat 5 Committee (1987)). The non-linear parameter k was estimated using a simple grid search over a range of values of k.

We now turn to the problem of duplicates. The CMIB offered two solutions to this problem. The first solution was to use the normal approximation to the Poisson distribution for the number of deaths; this approach may run into difficulties when the number of deaths is small and the normal approximation is poor. The second solution was to use minimum χ^2; this approach allows the estimation of the parameters but not of their standard errors. We took a third option and used the method of quasi-likelihood (McCullagh & Nelder, Chapter 9 (1989)). We know that without duplicates inference for the model in (4.1) and (4.2) can be expressed in terms of a Poisson log-likelihood; in the presence of duplicates we know both the mean and variance of the number of deaths. The method of quasi-likelihood allows us to combine both of these pieces of information in a single unified model for all durations.

In the simple Poisson case, if $Y \sim \mathrm{P}(\lambda)$ then $f(y) = e^{-\lambda}\lambda^y/y!$. The log-likelihood is:

$$\ell(\lambda) = y \log \lambda - \lambda.$$

This Poisson log-likelihood can be extended to a quasi-likelihood by taking:

$$\ell(\lambda) = w(y \log \lambda - \lambda)$$

where w is a weight function to be chosen. Clearly, if $w = 1$ we have the usual Poisson likelihood, while if $w \neq 1$ we have a variable with mean λ and variance λ/w. The appropriate weight function is $w = 1/r$, where r is the variance inflation factor for the age in question. Let θ, E and r be the number of claims, the central exposure and the variance inflation factor respectively, then the contribution to the log-likelihood at this age and duration is:

$$\ell(\mu) = (\theta \log \mu - E\mu)/r. \tag{4.3}$$

This gives $\hat{\mu} = \theta/E$ with estimated variance $r\theta/E^2$. Asymptotically, this is the same solution as that obtained using a normal approximation or a minimum χ^2 approach, while if $w = 1$ the exact log-likelihood is used. From the computational point of view this approach has one major advantage: the resulting model is still a generalised linear model and can be fitted in a straightforward way.

5. THE FITTED MODEL

In this section we report on the results of fitting the models, with systematic part given by (4.1) and random part given by the quasi-likelihood of (4.3). The models were fitted using data with ages 20 through 90 and curtate durations 0, 1, 2, 3, 4, 5+; duration 5+ used estimated mean durations for each age. We omitted the data for ages less than 20 years and greater than 90 years; in the case of ages less than 20 years there were too little data while the data for high ages were felt to be too unreliable.

The generalised linear model form of (4.1) and (4.3) made experimenting with the model straightforward. For example, the model allowed us to include an interaction between age and duration. We were aware that a weakness of our method of allocating all the data for duration 5+ to a single (estimated) duration may have made it less likely that an interaction would be found. There were three methods used to investigate the presence of an interaction between age and duration:

(a) informal methods based on grouping of data by age within duration and plotting the resulting raw death rates,
(b) inclusion of simple interaction terms via $f_3^*(x,d)$ in the model defined by (4.1) and (4.3), and
(c) fitting separate GM(0,s) models for each duration, and comparing the results with the model with the same age effect for all durations (see (5.2)).

The informal methods in (a) suggested that any interaction was weak, although it is possible that the evidence was diluted by the grouping. In the investigation of (b) we tried the following forms for $f_3^*(x,d)$:

$$\frac{x}{(d+k)}, \quad \frac{x^2}{(d+k)}, \quad \frac{x}{(d+k)^2}, \quad \frac{x^2}{(d+k)^2} \tag{5.1}$$

for suitably chosen k. None of these functions gave a significant improvment on the overall fit nor made any obvious improvement to the residual plots. We concluded that any interaction was not of the simple mixed polynomial type of (5.1). It is not easy to see what other continuous form for the interaction could be considered. The results for (c), which are discussed later in the section, confirmed the conclusion that any interaction was not of the simple form (5.1). The conclusion that we came to was that we could drop the interaction term, $f_3^*(x,d)$, and, more importantly, that a model without interaction could be found that gave an acceptable fit.

We were left with a multiplicative model for $\mu_{x,d}$ that consisted of two factors: one factor, $f_1^*(x)$, accounted for the dependence of $\mu_{x,d}$ on x, and the other factor, $f_2^*(d)$, accounted for the dependence of $\mu_{x,d}$ on d. An important property of this model is that the ratio of the forces of mortality at any age x does not depend on x, i.e.

$$\frac{\mu_{x,d_1}}{\mu_{x,d_2}} \quad \text{is free of } x.$$

Before reporting our results in detail we comment briefly on the choice of the form of the function $f_2^*(d)$. Recall that we wanted a function that was monotonic increasing in d and tended to a finite limit as $d \to \infty$. One possibility was to use a polynomial in $1/d$. We did try this approach, and while the model gave satisfactory results for large values of d it was very unstable for small d. We found that using a polynomial in $1/(d+k)$, where k was a further parameter, stabilised the fit.

Our final fitted model used polynomials of degree 6 in x and degree 3 in $1/(d+1{\cdot}28)$. The fitted parameter values, together with standard errors and t-values are given in Table 3. The fitted model gives a deviance ($-2 \times$ log-likelihood ratio) of 465·2 with 406 degrees of freedom, and thus a mean deviance

Table 3. *Estimated coefficients, standard errors and* t*-values*

	Coefficient	Standard error	t-value
a_0	−3·7990	0·09	−40·1
a_1	4·6043	0·19	24·6
a_2	5·1782	1·10	4·7
a_3	9·7293	2·17	4·5
a_4	−22·8717	6·61	−3·5
a_5	−58·6521	13·30	−4·4
a_6	−30·6477	7·28	−4·2
b_0	−6·7280	1·47	−4·6
b_1	21·9475	5·81	3·8
b_2	−23·9322	6·32	−3·8
k	1·28	1·5	0·8

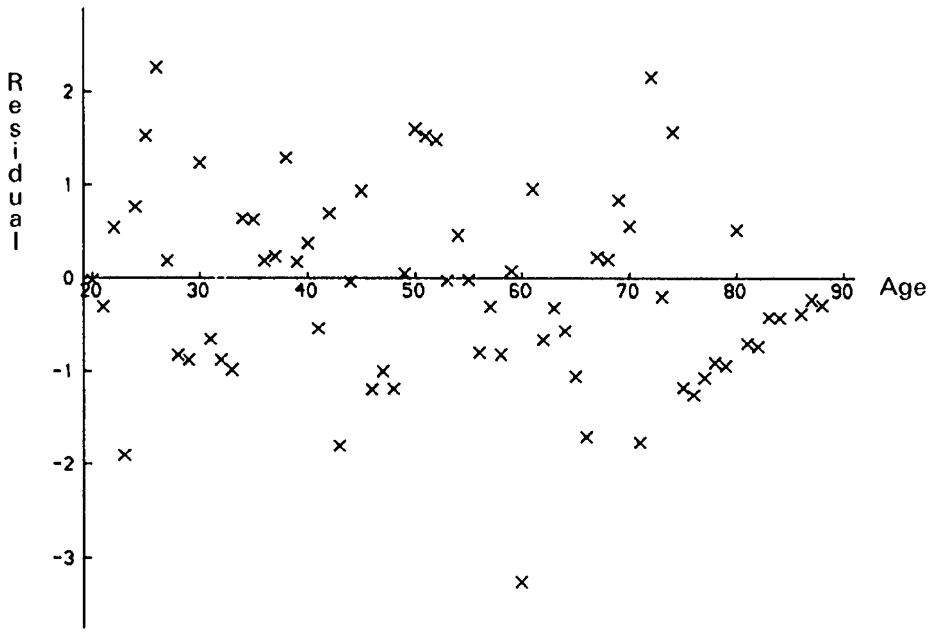

Figure 2. Residual plot for duration 0.

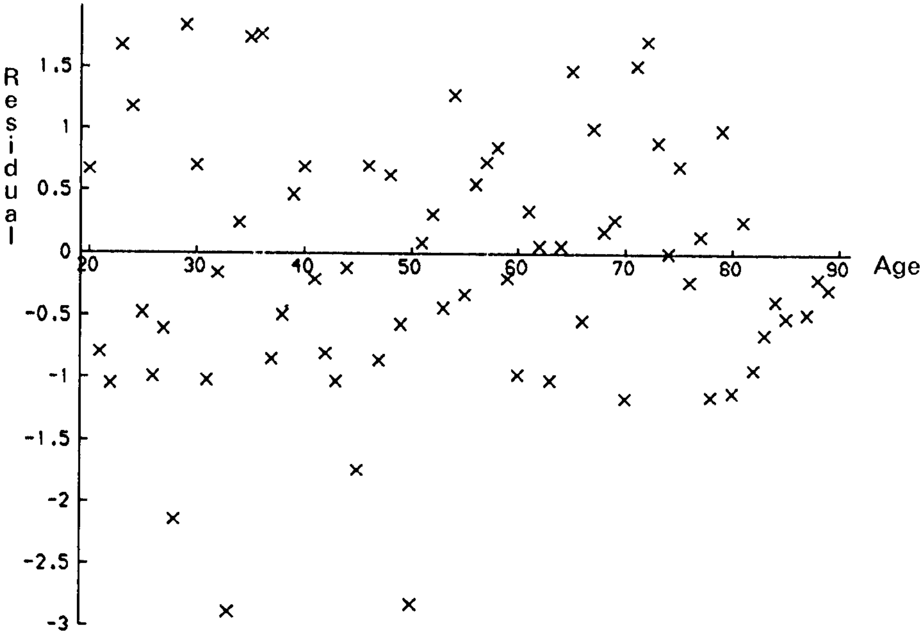

Figure 3. Residual plot for duration 1.

of 1·15. At first sight this figure may seem rather high (if the model is exactly true the mean deviance should be about 1). One possible explanation is that the estimates of the variance ratios are generally too low; this would certainly have the effect of inflating the residual variance. Another explanation is that the data are collected from a large number of different companies with different mixes of business and policyholders; the resulting heterogeneity in the data could also result in inflating the variance.

The deviance is a measure of over-all fit. A more detailed examination of the fit of the model can be made using the residuals, and Figures 2–7 give the residual plots for each duration separately. Examination of these plots does not reveal any obvious departure from the model; it is more that the residuals are generally rather large. However, one apparently unsatisfactory feature of the residual plots does stand out. For durations 0, 1, 2, 3 and 4 there is an obvious pattern to the residuals at high ages. Nearly all the residuals are negative and show a definite rising pattern. Examination of the original data reveals that these negative residuals result from very low exposures at these ages which give rise to zero deaths; the fitted number of deaths is, of course, positive.

Table 4 gives the deviances (asymptotically equivalent to the χ^2 values quoted

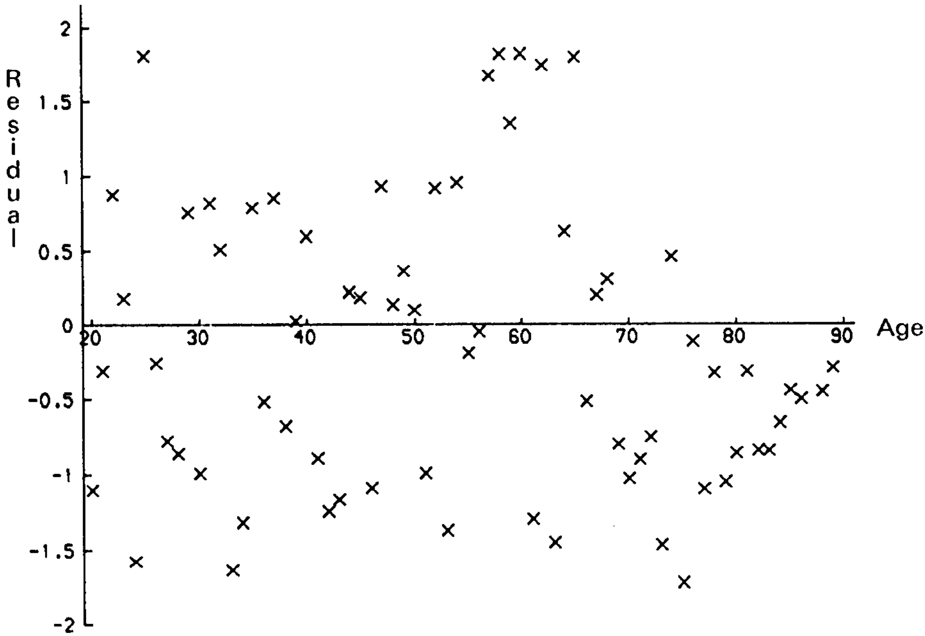

Figure 4. Residual plot for duration 2.

by the CMIB) for four different models and allows a general comparison between models to be made. Model 1 is a GM(2,2) model fitted separately for each duration. Thus 24 parameters are used in total. The overall fit is rather poor, mainly because the model fails to fit the duration 5+ data at all well. (The CMIB gives a χ^2 value for the GM(1, 5) model of 76·2.) Our joint model, Model 3, for age and duration defined by (4·1), (4·3) with coefficients in Table 3, gives very similar deviances to Model 1 with the exception of duration 5+. The total number of parameters fitted is 11. We were reassured that our joint model was doing as well as, if not better than, the CMIB's 'best-buy' model.

Our Model 3 can be regarded as a GM(0,7) model in age with adjustments depending on duration. We discuss two further GM (0,7) age effect models. We have included for reference the deviances for the GM (0,7) function (Model 4) fitted separately for each duration. The deviance of this model is 52·2 lower than our model, which has 31 fewer parameters. The remaining model included in Table 4 is what we loosely refer to as a factor model (Model 2). Model 2 is defined by:

$$\mu_{x,d} = \exp\{f_d + p(x)\}, \qquad \text{for } d = 0, 1, 2, 3, 4 \text{ and } 5+ \tag{5.2}$$

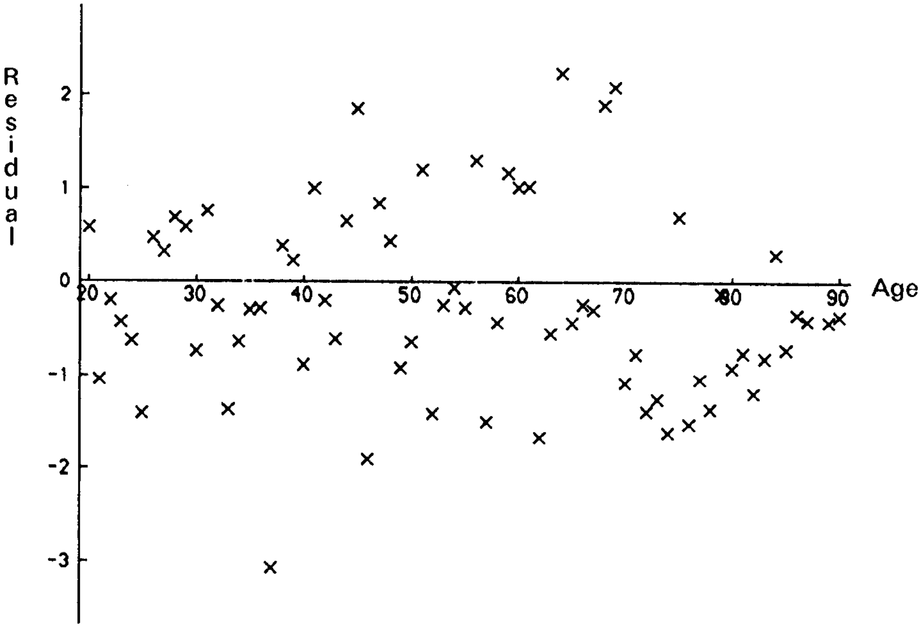

Figure 5. Residual plot for duration 3.

where $p(x)$ is a polynomial of degree 6 (in this case) and f_d are constants with $f_0 = 0$. Model 2 can be interpreted as a GM (0,7) model with no interaction fitted to all durations simultaneously. Model 2 is nested in (i.e. is a sub-model of) Model 4 and hence we can carry out a test for interaction between age and duration, as indicated in (c) above. The difference in deviance between Model 2 and Model 4 is 55·3 with 30 degrees of freedom, which is significant at $\frac{1}{2}$%. There is thus some evidence of interaction (see our comments at the beginning of this section). We make three comments on this finding. First, the interaction is not picked up by the product terms of (5.1). Second, the difference in deviance of 55·3 is partly explained by the fact that, unlike the CMIB, we have not grouped the data when small expected values occur. Third, the overall fit of Model 3, compared to the CMIB Model 1, as measured by the deviance of 465·2 and on the evidence of the residual plots, is quite good.

Our final comment on the fit of our model arises by comparing Model 2 and Model 3. Both these models describe the effect of duration by multiplying the duration 0 (base-line) mortality by a constant. Table 5 gives the factors by which mortality is inflated using the factor model and using the joint age and duration model. It is readily seen that there is good agreement between the two sets of

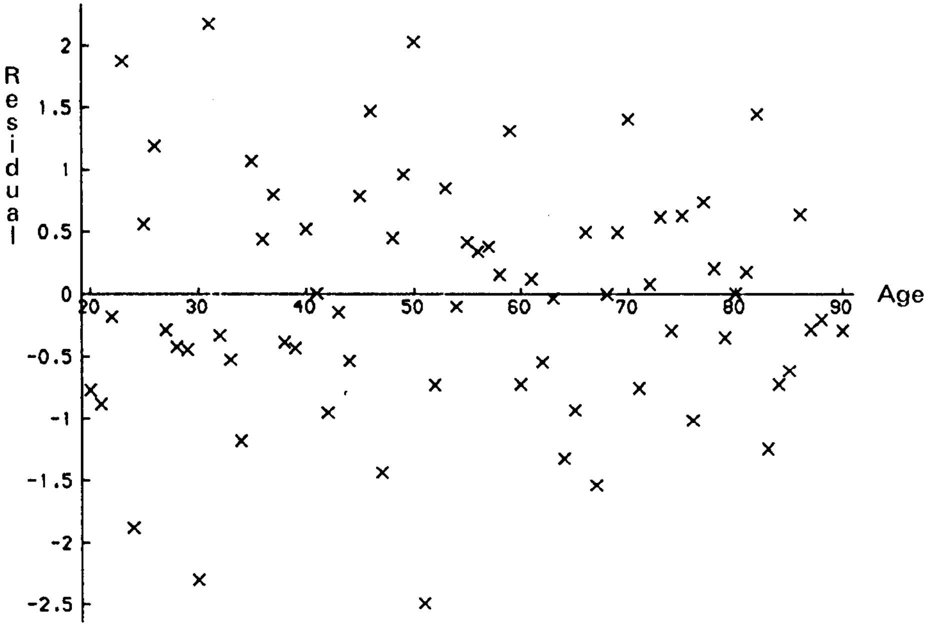

Figure 6. Residual plot for duration 4.

figures. The absence of an entry for duration 5+ under Model 3 underlines the difference between our approach and the other three approaches. Models 1, 2 and 4 treat the 5+ data as a single duration; in Model 3 we have attempted to treat duration in 5+ as a continuous variable. One possible way of filling the blank in Table 5 is to ask, under Model 3, what duration d gives a mortality ratio of 1·936. We easily find that d is about 10 since $\exp\{f_2^*(10)-f_2^*(0{\cdot}5)\}=1{\cdot}926$.

The multiplicative form of the model enables us to assess how well it fits the data separately for age and for duration since selection. Consider first a particular curtate duration, d. Let x be any (integer) age which satisfies the condition:

$$E(d,x)\mu_{x,d+0\cdot5}\geqslant 5. \tag{5.3}$$

(This condition ensures that the age/duration combination being considered has sufficient data for the normal approximation to the Poisson distribution to be reasonable. This in turn implies that point estimates will be approximately symmetrically distributed about their mean values.) Hence, approximately:

$$\theta(d,x)\sim \mathrm{N}(\exp\{f_1^*(x)\}\exp\{f_2^*(d+0{\cdot}5)\}E(d,x),\ r_x\theta(d,x))$$

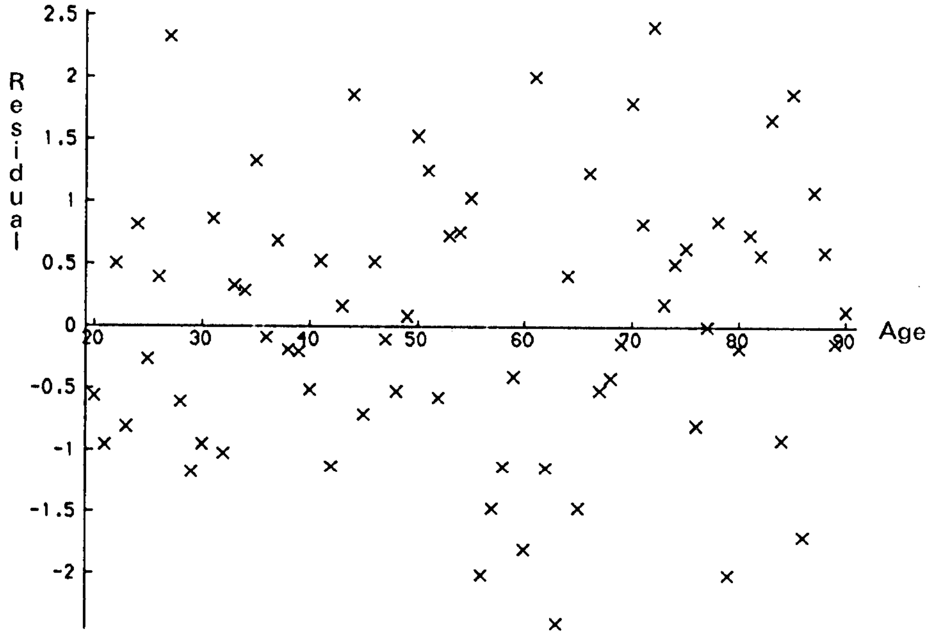

Figure 7. Residual plot for duration 5+.

from which we have:

$$\frac{\theta(d,x)}{\exp\{f_1^*(x)\}E(d,x)} \sim \mathrm{N}\left(\exp\{f_2^*(d+0\cdot5)\}, \frac{r_x\theta(d,x)}{(\exp\{f_1^*(x)\}E(d,x))^2}\right)$$

and finally:

$$\frac{1}{m}\sum_x \frac{\theta(d,x)}{\exp\{f_1^*(x)\}E(d,x)}$$

$$\sim \mathrm{N}\left(\exp\{f_2^*(d+0\cdot5)\}, \frac{1}{m^2}\sum_x \frac{r_x\theta(d,x)}{(\exp\{f_1^*(x)\}E(d,x))^2}\right) \quad (5.4)$$

where the summation is over all ages x satisfying (5.3) for the given duration d, and m is the number of ages included in each summation. Figure 8 shows a graph of $\exp\{f_2^*(d)\}$ for $\frac{1}{2} \leqslant d \leqslant 50$, together with point estimates:

Table 4. *Deviances and degrees of freedom for various models*

	Model 1		Model 2		Model 3		Model 4	
Duration	Deviance	Degrees of Freedom	Deviance	Degrees of Freedom	Deviance	Degrees of Freedom	Deviance	Degrees of Freedom
0	79·7	64	81·8	—	80·3	—	70·7	61
1	71·3	65	74·7	—	74·3	—	67·8	62
2	70·5	65	73·5	—	72·1	—	54·7	62
3	81·5	66	87·6	—	86·0	—	74·2	63
4	75·6	66	73·7	—	73·7	—	72·2	63
5+	107·1	67	77·0	—	78·8	—	73·4	64
Total	485·7	393	468·3	405	465·2	406	413·0	375

Model 1: GM(2,2) model fitted separately for each duration.
Model 2: GM(0,7) factor model.
Model 3: Model defined by (6.3).
Model 4: GM (0,7) model fitted separately for each duration.

Table 5. *The ratios* $\mu_{x,d+0\cdot5}/\mu_{x,0\cdot5}$, *for* d = *1, 2, 3, 4 and 5+*

Duration	Model 2	Model 3
1	1·495	1·495
2	1·514	1·506
3	1·495	1·538
4	1·615	1·592
5+	1·936	—

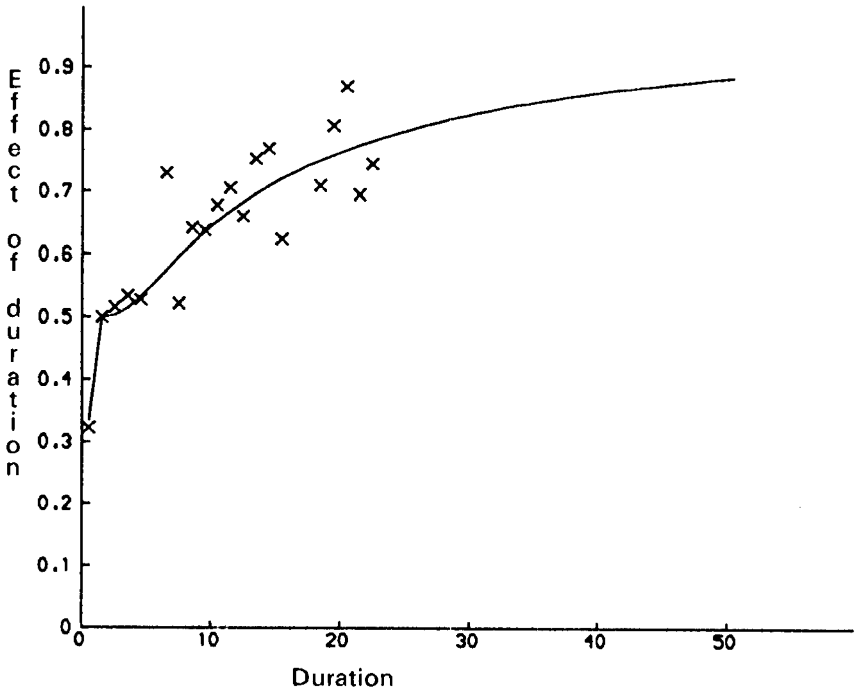

Figure 8. The effect of selection, $f_2^*(d)$.

$$\frac{1}{m}\sum_x \frac{0(d,x)}{\exp\{f_1^*(x)\}E(d,x)} \tag{5.5}$$

of $\exp\{f_2^*(d+0\cdot5)\}$ for $d = 0, 1, 2, \ldots, 22$. Figure 9 gives a close-up of the graph of $\exp\{f_2^*(d)\}$ together with standard deviations above and below the point estimates given by (5.4); in detail, Figure 9 gives the graph of $\exp\{f_2^*(d)\}$ for

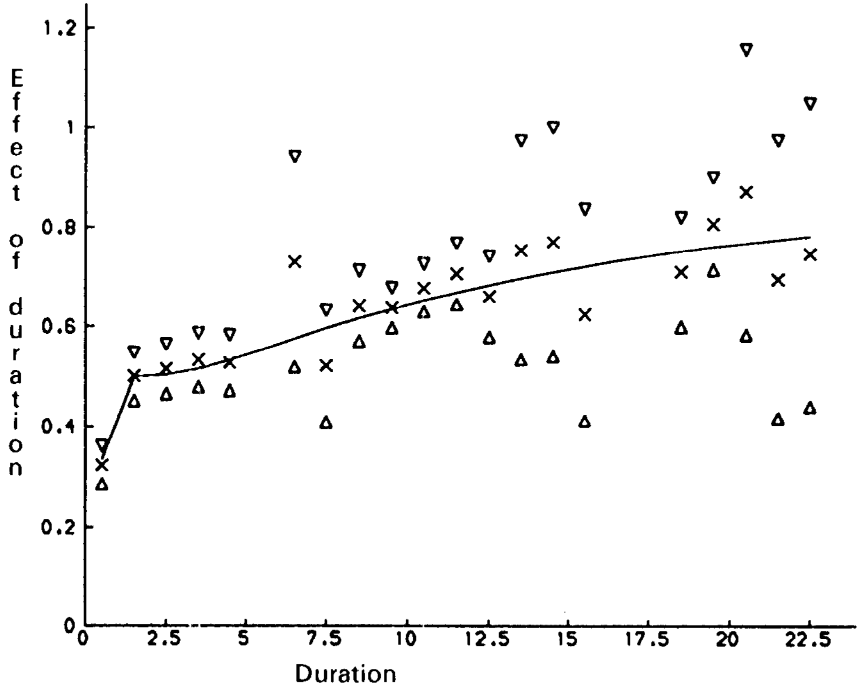

Figure 9. The effect of selection, $f_2^*(d)$.

$\frac{1}{2} \leqslant d \leqslant 23$, the point estimates of $\exp\{f_2^*(d+0{\cdot}5)\}$ and the 2 standard deviation bounds for these point estimates given by:

$$\frac{1}{m}\sum_x \frac{\theta(d,x)}{\exp\{f_1^*(x)\}E(d,x)} \pm \frac{2}{m}\left\{\sum_x \frac{r_x\theta(d,x)}{(\exp\{f_1^*(x)\}E(d,x))^2}\right\}^{\frac{1}{2}}$$

It can be seen from Figure 9 that the function $\exp\{f_2^*(d)\}$ fits the data very well; for all 23 durations the function is within 2 standard deviations of the point estimate. The graphs of $\exp\{f_2^*(d)\}$ illustrate the effect on the force of mortality of duration since selection. This effect will be discussed further in Section 7.

For any given (integer) age x, the expression corresponding to (5.4) is:

$$\frac{1}{n}\sum_d \frac{\theta(d,x)}{\exp\{f_2^*(d+0{\cdot}5)\}E(d,x)}$$

$$\sim \mathrm{N}\left(\exp\{f_1^*(x)\}, \frac{1}{n^2}\sum_d \frac{r_x\theta(d,x)}{(\exp\{f_2^*(d+0{\cdot}5)\}E(d,x))^2}\right) \quad (5.6)$$

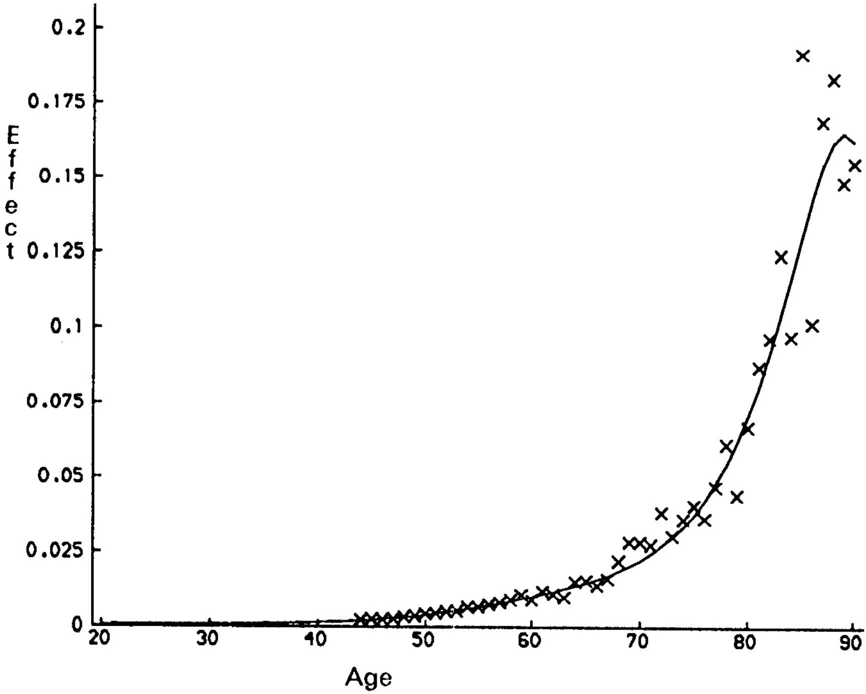

Figure 10. The effect of age, $f_1^*(x)$.

where the summations are now over all curtate durations d satisfying condition (5.3) for this particular x, and n is the number of terms in each summation. Figures 10 and 11 show the graphs of $\exp\{f_1^*(x)\}$ for $20 \leqslant x \leqslant 90$ and $20 \leqslant x \leqslant 55$, respectively, together with the values of:

$$\frac{1}{n}\sum_d \frac{\theta(d,x)}{\exp\{f_2^*(d+0{\cdot}5)\}E(d,x)}$$

for integer values of x, which, from (5.6), can be regarded as a point estimate of $\exp\{f_1^*(x)\}$. In this case we have shown separately the graph of $\exp\{f_1^*(x)\}$ together with the point estimates for $20 \leqslant x \leqslant 55$, (i.e. Figure 11), in order to show more clearly what is happening in this age range. It would have been possible, using (5.6), to show in Figures 10 and 11 the point estimates $\pm$ 2 standard deviations, as we did in Figure 9. We have not done so because we did not wish to obscure the more important points in these Figures. However, it can be seen from Figures 10 and 11 that the function $\exp\{f_1^*(x)\}$ fits the point estimates very well. In fact, there are only six ages, out of a total of the 68 for

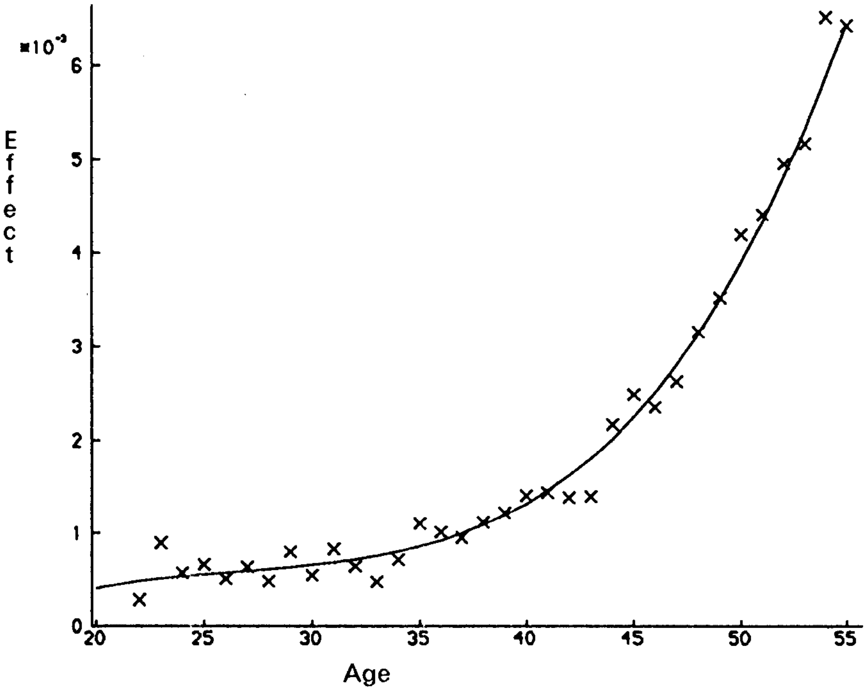

Figure 11. The effect of age, $f_1^*(x)$.

which we have data, where the graduated value of $\exp\{f_1^*(x)\}$ is more than 2 standard deviations away from the corresponding point estimate. (Two of these are adjacent ages, 85 and 86, where the point estimates are above and below the graduated value of $\exp\{f_1^*(x)\}$, respectively.) It can be seen from Figure 10 that $\exp\{f_1^*(x)\}$ is decreasing for $x > 89$. This point will be discussed further in the following section.

6. ADJUSTMENTS AT HIGH AGES AND SHORT DURATIONS

It was shown in the previous section that the model for $\mu_{x,d}$ fitted the data very well within the ranges of age and duration represented in the data. However, when this model is extrapolated beyond those ranges it does not give acceptable values. In particular:

(a) for $x > 89$ the function $f_1^*(x)$ is decreasing, and

(b) $\lim_{d \to 0+} f_2^*(d)$ is too small.

We have corrected these two features by 'manually adjusting' $f_1^*(x)$ and $f_2^*(d)$. Note that the multiplicative form of the model makes it relatively simple to do this. (Another problem is that for $x < 20$, $f_1^*(x)$ gives values which are too small. We have not attempted to correct this feature.)

The adjustment at high ages was made as follows:

for some suitably high age x_0 (the value eventually chosen was $x_0 = 86$), define:

$$\left.\begin{array}{ll} f_1(x) = f_1^*(x) & \text{for } x < x_0 \\ f_1(x) = \alpha + \beta x & \text{for } x \geqslant x_0 \end{array}\right\} \tag{6.1}$$

where α and β are chosen so that $f_1(x)$ is continuous and has a continuous first derivative at x_0. Trials with various values for x_0 showed that $x_0 = 86$ gave the best extrapolated values of $f_1(x)$ for x up to age 110 and also gave the smoothest 'splice' at age x_0. Note that for $x \geqslant x_0$, $\exp\{f_1(x)\}$ is just a Gompertz formula.

The problem at short durations was that, although $\exp\{f_1^*(x) + f_2^*(\frac{1}{2})\}$ fitted the estimates $\theta_{x,0}/E_{x,0}$ reasonably well, the values of:

$$1 - \exp\left\{ - \int_{d=0}^{1} \exp[f_1^*(x+d) + f_2^*(d)]dd \right\}$$

did not fit the crude estimates of $q_{[x]}$, i.e. $\theta_{x,0}/(E_{x,0} + \frac{1}{2}\theta_{x,0})$, at all well, the latter generally being higher than the former. This problem was corrected as follows:

$$\left.\begin{array}{ll} f_2(d) = f_2^*(d) & \text{for } d > 1/2 \\ f_2(d) = \gamma + \delta/(d+k) & \text{for } d \leqslant 1/2 \end{array}\right\} \tag{6.2}$$

with γ and δ chosen so that $f_2(d)$ is continuous and has a continuous first derivative at $d = \frac{1}{2}$.

Hence, our final fitted model, which we regard as acceptable over the ranges $20 \leqslant x \leqslant 110$ and $0 \leqslant d < \infty$, is as follows:

$$\mu_{x,d} = \exp\{f_1(x) + f_2(d)\}$$

where for $x \leqslant 86$

$$f_1(x) = -3{\cdot}7990 + 4{\cdot}6043z + 5{\cdot}1782z^2 + 9{\cdot}7293z^3 - 22{\cdot}8717z^4 - 58{\cdot}6521z^5 - 30{\cdot}6477z^6$$

and where for $x > 86$

$$f_1(x) = -3{\cdot}2956 + 4{\cdot}2173z$$

where $z = (x - 70)/50$, and where for $d \geqslant \frac{1}{2}$ (6.3)

$$f_2(d) = \frac{-6{\cdot}7280}{d + 1{\cdot}28} + \frac{21{\cdot}9475}{(d+1{\cdot}28)^2} - \frac{23{\cdot}9322}{(d+1{\cdot}28)^3}$$

and where for $d < \frac{1}{2}$

$$f_2(d) = 1{\cdot}5600 - \frac{4{\cdot}7281}{d + 1{\cdot}28}.$$

7. SOME COMMENTS ON THE EFFECT OF DURATION SINCE SELECTION

For any attained age, $\exp\{f_2(d_2) - f_2(d_1)\}$ represents the ratio of the forces of mortality at durations d_2 and d_1. Hence the graph of $\exp\{f_2(d)\}$ shown (for $d \geqslant \frac{1}{2}$) in Figure 8 provides us with an indication of the effect on mortality of the duration since selection. In broad terms this graph confirms our not unreasonable preconceptions concerning this effect, namely that, for any attained age:

(a) the force of mortality increases with duration since selection, and
(b) the effect of selection becomes less as duration since selection increases, at least for durations in excess of about 5 years.

(Norberg (1988) gives a very succinct summary of our preconceptions concerning selection, and also gives a simple probabilistic model to explain them.)

However, some features of the function $\exp\{f_2(d)\}$ call for comment. The first is the 'kink' at the short durations. As a result of the CMIB's grouping of durations, this feature manifests itself in the original data as duration 0 having significantly lower, and duration 5 + having significantly higher mortality than durations 1, 2, 3 and 4, with no significant differences between these latter 4 durations (see the comments in *CMIR* **9** (1988, Section 6.2) and also Table 5). This feature seems to occur frequently in assured lives data; see, for example, the comments in *CMIR* **9** (1988, Section 3.3) concerning the 1979–1982 Permanent Assurances, Males, data and comments by the Joint Mortality Investigation Committee (1974, §3.4) concerning the 1967–70 Assured Lives data. This last reference also includes the statement, "The same features appeared in the 1949–1952 data and in earlier investigations . . .".

Another feature of the function $\exp\{f_2(d)\}$ apparent from Figure 8 and calling for comment, is the extrapolation of the function beyond about duration 20, which is where our data end. Figure 8 shows that although the effect of selection gradually wears off, as it is constrained to do from the choice of function for $f_2(d)$, it takes a very long time to do so! For example:

$$\exp\{f_2(30) - f_2(20)\} = 1{\cdot}055$$

so that for any attained age an individual selected 30 years ago is expected to have a force of mortality higher by $5\frac{1}{2}\%$ than an individual selected 20 years ago. This seems somewhat uncomfortable in the U.K. where we are used to having a select period of 2 or at most 5 years! However, two points can be made in support of our graduation:

(i) Assured lives data in the U.S.A. and Canada are collected and graduated with a 15-year select period (see, for example, the report of the Committee on Mortality under Ordinary Insurance and Annuities (1973) and Jones & Aitken (1990)).
(ii) The extrapolation of $\exp\{f_2(d)\}$ beyond duration 20 years, as shown in Figure 8, does not appear at all unreasonable.

A final point which should be made is that our data for high durations ($d > 13$), are based solely on the data for high ages ($x > 70$) and our data for lower durations ($5 < d < 8$) is based solely on data for young ages ($x < 40$) (see Figure 1). For this reason it is not easy to distinguish between the effects of 'temporary initial selection', i.e. selection as a result of passing a medical examination, and 'time selection', i.e. the effect of the mortality rate at any given age changing over time (see Benjamin & Pollard (1980, pp. 215–19)).

8. COMPARISONS WITH THE CMIB'S GRADUATIONS

It is of some interest to compare our graduation of the 1979–1982 Permanent Assurances, Females, data with the graduation of the same data by the CMIB. For this purpose we have taken the CMIB's graduation to be the AF80 table published in *CMIR* **10** (1990) (see Section 3 above). We shall denote the CMIB's graduations by $\tilde{\mu}_{x,0}$, $\tilde{\mu}_{x,1}$ and $\tilde{\mu}_{x,2+}$ in an obvious notation.

Table 6 shows ratios of the values of μ given by our graduation to the values given by the CMIB. More precisely, for each attained age x, i.e. for each row in the table, the figures shown are:

$$\frac{\mu_{x,0\cdot5}}{\tilde{\mu}_{x,0}}, \quad \frac{\mu_{x,1\cdot5}}{\tilde{\mu}_{x,1}}, \quad \frac{\mu_{x,D(x)}}{\tilde{\mu}_{x,2+}}.$$

Table 6. *Comparison of μ values: ratio of our values to the CMIB's*

Age	Duration 0·5	1·5	2+
20	0·590	0·749	0·790
30	1·010	1·074	1·115
40	0·906	0·989	0·988
50	1·011	1·059	1·134
60	1·065	0·980	1·065
70	1·083	0·831	0·849
80	1·463	1·014	1·135
90	1·627	1·127	1·212

Table 7. *Comparison of* q *values: ratios of our values to the CMIB's*

Attained age x_0+d	Entry age x_0 20	30	40	50	60	70
20	0·614					
30	1·205	0·966				
40	1·226	1·042	0·877			
50	1·453	1·349	1·147	0·984		
60	1·376	1·320	1·226	1·043	1·026	
70	1·159	1·129	1·083	1·006	0·856	1·055
80	1·367	1·343	1·309	1·256	1·169	0·998
90	1·371	1·354	1·331	1·299	1·250	1·168
100	1·163	1·153	1·141	1·124	1·100	1·063
110	1·017	1·012	1·006	0·998	0·987	0·972

Table 8. *Comparison of premium rates: ratio of our values to the CMIB's*

Term	Policy	Entry age 25	35	45	55
10	Endowment Assurance	1·000	0·999	0·999	0·996
	Temporary Assurance	1·040	0·922	0·993	0·909
20	Endowment Assurance	0·999	1·001	1·004	0·984
	Temporary Assurance	1·099	1·102	1·074	0·935
30	Endowment Assurance	1·007	1·015	1·017	0·984
	Temporary Assurance	1·211	1·169	1·067	1·015
—	Whole Life Assurance	1·170	1·123	1·077	1·012

Table 6 does not reveal any consistent pattern for these ratios being greater than or less than 1. Where the data are most numerous, i.e. durations 2 + (all ages) and ages 50 and 60 (select durations), the ratios are generally very close to 1.

Table 7 shows ratios of values of $q_{[x_0]+d}$ given by our graduation to the values given by the CMIB for various entry ages x_0, and various attained ages $[x_0]+d$. Compared to Table 6, Table 7 shows a much higher proportion of ratios greater than 1 and also some relatively high individual ratios, e.g. $q_{[20]+30}/\tilde{q}_{[20]+30}$. The high value for this particular ratio may seem surprising, especially when from Table 6 we see that $\mu_{50,D(50)}/\tilde{\mu}_{50,2+} = 1{\cdot}134$. However, it should be recalled that $D(50) = 10{\cdot}17$ and that $\exp\{f_2(30) - f_2(10{\cdot}17)\} = 1{\cdot}275$.

Table 8 shows ratios of net premium rates given by our graduation to those given by the CMIB's graduations. The policies are endowment assurances and term assurances for terms up to 30 years, together with whole life assurances, all for various entry ages up to age 55. In all cases the rate of interest is 8% p.a., the sum assured is payable immediately on death (or survival to the end of the term) and premiums are payable continuously at a level rate throughout the term of the policy. Perhaps not surprisingly, the ratios in Table 8 age generally closer to 1 than those in Table 7.

9. CONCLUDING REMARKS

Our purpose in this paper has been to present an approach to the graduation of select mortality data jointly by age and duration, and to illustrate this approach by graduating the data for 1979–82, Permanent Assurances, Females. In doing so we encountered a number of difficulties. The most notable of these have been:

(i) Having to estimate the function $D(x)$, i.e. the average duration for policies whose duration exceeds 5 years. The estimation of $D(x)$ has been a key element in our graduation and it was unfortunate that a more precise method for its estimation than the method detailed in our Appendix was not available. Nevertheless, we regard our estimates of $D(x)$ as not unreasonable.

(ii) Interaction between age and duration since selection. As detailed in Section 5, there was some evidence in the data of interaction between age and duration since selection which terms of the form (5.1) we were unable to model.

(iii) Time selection. We have attempted to model temporary initial selection and our final model indicates that the effect of temporary initial selection could last a long (strictly speaking an infinite!) time. In these circumstances it is difficult, without carrying out a more detailed study, to separate the effects of temporary initial selection from time selection. However, we make no apologies for this difficulty since it is a consequence of the nature of the data available. Our method of graduation at least has the merit of highlighting rather than disguising the problem.

Despite the problems mentioned above, we believe the present study to be interesting and useful for the reasons mentioned in Section 1. A final small, but interesting, point of comparison between our graduation and the CMIB's is that if we consider an individual selected at age x_0 and then consider the force of mortality at some duration d ($\geqslant 0$), our graduation gives a force of mortality, $\mu_{x_0+d,d}$, which is a continuous (and differentiable) function of d, whereas the CMIB's graduations give a force of mortality, $\tilde{\mu}_{x_0+d,0}$ for $0 \leqslant d < 1$, $\tilde{\mu}_{x_0+d,1}$ for $1 \leqslant d < 2$, etc., which has points of discontinuity.

REFERENCES

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

CMIR **9** (1988). Report No. 9 of the Continuous Mortality Investigation Bureau. Faculty of Actuaries and Institute of Actuaries.

CMIR **10** (1990). Report No. 10 of the Continuous Mortality Investigation Bureau. Faculty of Actuaries and Institute of Actuaries.

Currie, I. D. (1990). Discussion of 'On graduation by mathematical formula' by D. O. Forfar, J. J. McCutcheon & A. D. Wilkie. *T.F.A.* **41**, 97–269.

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

GENSTAT 5 COMMITTEE. (1987). *Genstat 5: Reference Manual.* Clarendon Press, Oxford.

JOINT MORTALITY INVESTIGATION COMMITTEE (1974). Considerations affecting the Preparation of Standard Tables of Mortality. *J.I.A.* **101**, 133–201.

JONES, B. L. & AITKEN, W. H. (1990). Mortality experience of Canadian insured lives during 1982-1986. Research Report 90 08 published by the Institute of Insurance and Pension Research, University of Waterloo, Canada.

LYONS, T. J. (1990). Discussion of 'On graduation by mathematical formula' by D. O. Forfar, J. J. McCutcheon & A. D. Wilkie. *T.F.A.* **41**, 97–269.

MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized Linear Models, Second Edition.* Chapman and Hall, London.

NELDER, J. A. & WEDDERBURN, R. W. M. (1972). Generalized Linear Models. *J.R. Statist. Soc., A*, **135**, 370–84.

NORBERG, R. (1988). Select mortality: possible explanations. *Transactions of the International Congress of Actuaries*, **3**, 215–23.

PANJER, H. H. & RUSSO, G. (1990). Parametric graduation of Canadian individual mortality experience: 1982–1988. Research Report 90–13 published by the Institute of Insurance and Pension Research, University of Waterloo, Canada.

RENSHAW, A. E. (1991). Actuarial graduation practice and generalised linear and non-linear models. *J.I.A.* **118**, 295–312.

REPORT OF THE COMMITTEE ON MORTALITY UNDER ORDINARY INSURANCE AND ANNUITIES (1973). 1965–1970 Basic tables. *Transactions of the Society of Actuaries, 1973 Reports*, 199–223.

TENENBEIN, A. & VANDERHOOF, I. T. (1980). New mathematical laws of select and ultimate mortality. *Trans. Soc. Act.* **XXXII**, 119–158.

APPENDIX

THE ESTIMATION OF $D(x)$

In this Appendix we give details of the method we have used to estimate the function $D(x)$, the average duration of policy for policyholders aged x nearest birthday who effected their policies 5 or more years ago.

In outline, our method is to estimate for each age x and for $d \geqslant 5$ the value of $E(d,x)$. Then, noting that:

$$E(5+,x) = \sum_{d \geqslant 5} E(d,x) \tag{A.1}$$

we can estimate $D(x)$ from the formula:

$$D(x) = \left\{ \sum_{d \geqslant 5} (d + \tfrac{1}{2}) E(d,x) \right\} / E(5+,x). \tag{A.2}$$

(We have assumed that the individual terms in the summation in (A.1) do not contain duplicate policies, even though $E(5+,x)$ does contain duplicates. The reason for this is that we assume that duplicates generally arise from policies affected at different times rather than at the same time. If this were not the case, we could have been expected to have found evidence of duplicates in the data for policies of the same duration with the duration less than 5 years, which we did not (see *CMIR* **9**, Section 6.1).)

Let us consider the estimation of $E(d,x)$ for a given x and a given $d \geqslant 5$. To do this we need some new notation. We denote by:

$E(k, y, m)$ the exposure from age $(y - \frac{1}{2})$ to age $(y + \frac{1}{2})$ and from duration k to duration $(k + 1)$ if the observation period had been $(1979 - m)$ to $(1982 - m)$, i.e. if the observation period had been m years earlier than the actual observation period,

$w_{[y]+t}$ the probability of withdrawing (i.e. of the policy lapsing or maturing) within 1 year for a policyholder aged $(y + t)$ who effected her policy t years ago,

$g(y)$ the annual rate of increase in the number of policies effected at age y. (We assume $g(y)$ is constant over time for each y.)

We can now write down the following (approximate) relationships:

$$E(d,x) \approx E(4,x - d + 4,d - 4) \times \prod_{t=4}^{d-1} \{(1 - q_{[x-d-\frac{1}{2}]+t+\frac{1}{2}})(1 - w_{[x-d-\frac{1}{2}]+t+\frac{1}{2}})\} \tag{A.3}$$

$$E(4,x - d + 4, d - 4) \approx E(4,x - d + 4)/(1 + g(x - d - \tfrac{1}{2}))^{d-4}. \tag{A.4}$$

Loosely speaking, formula (A.3) says that the difference between $E(4,x-d+4,d-4)$ and $E(d,x)$ is that some policyholders who contribute to the former may die or withdraw before contributing to the latter. Formula (A.4) says that the difference between $E(4,x-d+4)$ and $E(4,x-d+4,d-4)$ results from the increase in the number of policies issued at age $(x-d-\frac{1}{2})$. Values of $E(4,y)$ for $y \geqslant 10$ are available to us, so the problem of estimating $E(d,x)$ has become a problem of estimating $q_{[x-d-\frac{1}{2}]+t+\frac{1}{2}}$, $w_{[x-d-\frac{1}{2}]+t+\frac{1}{2}}$ and $g(x-d-\frac{1}{2})$. For $t \geqslant 4$ we have estimated $q_{[x-d-\frac{1}{2}]+t+\frac{1}{2}}$ simply from the formula:

$$q_{[x-d-\frac{1}{2}]+t+\frac{1}{2}} \approx \theta(5+,x-d+t)/E(5+,x-d+t). \tag{A.5}$$

Formula (A.5) gives a rather crude estimate of $q_{[x-d-\frac{1}{2}]+t+\frac{1}{2}}$. The estimates depend on attained age but are independent of duration since the policy was effected (for durations in excess of $4\frac{1}{2}$ years). The relative error introduced here should be small, and the relative error in the estimate of $(1-q_{[x-d-\frac{1}{2}]+t+\frac{1}{2}})$, which is the factor in which we are interested, should be even smaller since $q_{[x-d-\frac{1}{2}]+t+\frac{1}{2}}$ itself is small.

We have estimated $g(y-\frac{1}{2})$ from the formulae:

$$(1+g(y-\tfrac{1}{2})) = \left\{\prod_{t=-2}^{2}(1+g^*(y-\tfrac{1}{2}+t))\right\}^{\frac{1}{5}} \tag{A.6}$$

$$(1+g^*(y-\tfrac{1}{2})) \approx E(3,y+3)/E(3,y+3,1) \tag{A.7}$$

$$E(3,y+3,1) \approx E(4,y+4)/\{(1-q_{[y-\frac{1}{2}]+3\frac{1}{2}})(1-w_{[y-\frac{1}{2}]+3\frac{1}{2}})\}. \tag{A.8}$$

Formula (A.8) corresponds to formula (A.3) and formula (A.7) corresponds to formula (A.4). We regard $g^*(y-\frac{1}{2})$ as a crude estimate of $g(y-\frac{1}{2})$ and have calculated the latter from the former by taking a geometric mean of five consecutive values (formula (A.6)).

We have estimated $q_{[y-\frac{1}{2}]+3\frac{1}{2}}$ by:

$$q_{[y-\frac{1}{2}]+3\frac{1}{2}} \approx \tfrac{1}{2}\left\{\frac{\theta(3,y+3)}{E(3,y+3)} + \frac{\theta(4,y+3)}{E(4,y+3)}\right\}. \tag{A.9}$$

The remaining step in the estimation of $E(d,x)$, and hence of $D(x)$, is the estimation of $w_{[y-\frac{1}{2}]+t+\frac{1}{2}}$ for $t \geqslant 3$. We have made the following assumption:

$$w_{[y-\frac{1}{2}]+t+\frac{1}{2}} \quad \text{is a constant, independent of } y \text{ and of } t \geqslant 3. \tag{A.10}$$

It is easily seen that the actual value of this constant does not affect the estimated value of $E(d,x)$.

In practice, the estimated values of $E(d,x)$ for $d \geqslant 5$ will not sum exactly to (the known value of) $E(5+,x)$. In general, we found that the summation exceeded the value of $E(5+,x)$. In such cases, we truncated the summation at the duration $d(x)$ where:

$$E(5+,x) = \sum_{d=5}^{d(x)} E(d,x) \tag{A.11}$$

(taking only that part of $E(d(x),x)$ necessary to give equality in (A.11)) and then estimated $D(x)$ using:

$$D(x) \approx \left\{ \sum_{d=5}^{d(x)} (d + \tfrac{1}{2}) E(d,x) \right\} / E(5+,x). \tag{A.12}$$

Our justification for this procedure is that we have more confidence in our method for estimating $E(d,x)$ for lower values of d. In cases where the summation in (A.1) gave a value less than $E(5+,x)$, which occurred for a few of the higher ages, we estimated $D(x)$ using:

$$D(x) \approx \left\{ \sum_{d \geqslant 5} (d + \tfrac{1}{2}) E(d,x) \right\} \Big/ \left\{ \sum_{d \geqslant 5} E(d,x) \right\}.$$