JIA 114 (1987) 339-369

GRADUATION: SOME EXPERIMENTS WITH KERNEL METHODS

By D. S. F. Bloomfield, B.Sc. and
Professor S. Haberman, M.A., Ph.D., F.I.A.

1. INTRODUCTION

1.1 *Graduation: methods and uses*

It may be useful to begin with some introductory comments on the nature of graduation and its uses with particular reference to the distinction between graduation by parametric methods, as described in traditional actuarial textbooks, and graduation by non-parametric methods.

Some of the comments have appeared in the early sections of Copas and Haberman's paper which first described the kernel method of non-parametric graduation.[(1)]

Graduation may be regarded as the principles by which a set of data is adjusted in order to provide a basis suitable for inferences to be drawn and further practical calculations to be made. In actuarial terminology, graduation usually refers to a set of decremental rates, and one of the principal actuarial examples of its use is the construction of a life table from a set of age-specific, observed death rates.

The fundamental justification for the graduation of a set of observed data, say mortality or morbidity rates specific for age x, is the premise (supported by empirical evidence) that, if the number of units in the group on whose experience the data are based, n_x, had been considerably larger, the set of observed data (here, rates) would have a much more regular (i.e. smoother) progression with x (here, age). In the limit, with n_x indefinitely large, the set of data would have exhibited a smooth progression with x.

Thus the observed data may be regarded as a sample from a large population so that statistics derived therefrom are subject to sampling errors. Providing these errors are random in nature, they may be reduced by increasing the sample size i.e. enlarging the original investigation. A simpler, cheaper, quicker and more practicable alternative is often to use graduation to remove these random errors.

This discussion of the background to graduation has touched on the purposes for (and uses of) graduation. Thinking of the application to mortality and morbidity, it appears that there are five distinct uses of graduation. For completenesss, these are listed below:

(a) To smooth the data. Thus graduation facilitates the processing of the data, makes it easier to handle and removes awkward irregularities and inconsistencies. Both parametric and non-parametric methods of graduation have the potential to achieve smoothness but, as will be demonstrated in a later section, depending on how smoothness is defined, parametric methods may be superior in the degree of smoothness attained. There is both a practical and aesthetic side to smoothness. A set of premium rates for whole life insurances should look smooth and be smooth. A fortiori, if a reasonable degree of smoothness has not been achieved, complicated derived functions like policy values may display worrying irregularities.

(b) To make the results more precise, on the assumption that the true experience underlying the observations follows a smooth curve, as mentioned above. Both parametric and non-parametric methods can produce close adherence to the data, as will be discussed in a later section.

The following three uses are of less significance in an actuarial context:

(c) To aid inferences from incomplete data. In those populations for which complete registration of events, like births and deaths, is not available, indirect methods of estimation based on graduation are important.

(d) To facilitate comparisons of mortality. One would like to be able to compare the mortality of two populations, or of two cohorts or of one population at two points in time, summarizing the difference in a set of parameters. The parametric methods of graduation have the advantage of encapsulating the statistical information contained in the observations in a small number of parameters and hence enabling such comparisons to be easily made.

(e) To assist forecasting and projection. An important special example of comparisons is those over time. A clear progression over recent time in the values of a set of parameters enables extrapolation into the future to be used for forecasting of probabilities, rates and derived functions like life tables.

In (c), (d) and (e) the emphasis is clearly on parametric methods of graduation. This is less obvious for the more important uses, (a) and (b). It is in this context that a particular type of non-parametric graduation method will be described and applied.

Without loss of generality, in the description and discussion that follows we shall consider only the graduation of mortality rates q_x.

1.2 *Kernel method of non-parametric graduation*

A non-parametric method of graduation employing a kernel function, has been described and applied to a standard set of mortality rates by Copas and Haberman.[1] At each age x, the estimated probability of death $\hat{q}_x$, is given by the formula

$$\hat{q}_x = \frac{\sum_i s_i \psi(u_i)}{\sum_i n_i \psi(u_i)}$$

where n_i and s_i represent exposed to risk and recorded deaths respectively at age x_i, and $\psi(\ \ )$ is the kernel function. The variable u_i equals $(x-x_i)/h$, where h is a constant. u_i represents the distance from x to the data point x_i measured on a scale defined by h.

The normal kernel, $\psi_N(u)=e^{-\frac{1}{2}u^2}$, is used throughout this paper. Thus, in this work, h is the standard deviation of the distribution of the weights $\psi_N(u)$. Thus, in estimating $\hat{q}_x$, nearly all the information (i.e. over 95%) comes from observations in the range $x \pm 2h$.

If h is very small, the estimate is essentially the crude death rate at age x. As h increases, observations at other ages have greater influence on $\hat{q}_x$ and more smoothing occurs at the expense of the fit between graduated rates and the actual data. As $h \to \infty$, $\hat{q}_x$ tends to the overall probability of death. Thus, h measures in some way the information contributing to an estimate, $\hat{q}_x$.

2. BIAS OF THE KERNEL METHOD

Any graduation method will involve bias. The theory of exposed to risk and the principle of correspondence when strictly applied lead to crude mortality rates, $\mathring{q}_x$, which are unbiased estimators of the true mortality rates, q_x.

In order to make the crude mortality rates more smooth, we inevitably *introduce some bias or distortion via the process of graduation.* Bias is simply the price that we pay for this desirable feature of smoothness. It is usual to control the degree of bias by the use of statistical tests (mentioned briefly in section 3).

The bias of the kernel method of graduation is explored in the following paragraphs.

If we let the indicator random variable Z_i take the value 1 if death occurs at age x_i and 0 otherwise, then the estimate q_x may be rewritten:

$$\hat{q}_x = \frac{\sum_i Z_i \psi^{(h)}_{x,i} n_i}{\sum_i \psi^{(h)}_{x,i}\, n_i} \qquad \text{with} \quad \psi^{(h)}_{x,i} = \psi\left(\frac{x-x_i}{h}\right)$$

If we let q_{x_i} be the true probability of dying for age x_i, then the Z_i's are independent and each Z_i has a binomial distribution with mean q_{x_i} and variance $q_{x_i}(1-q_{x_i})$.

$$\text{Hence} \quad E(\hat{q}_x) = \frac{\sum_i E(Z_i) n_i \psi^{(h)}_{x,i}}{\sum_i n_i \psi^{(h)}_{x,i}}$$

$$= q_x + \text{BIAS} \quad \text{where BIAS} = \frac{\sum_i (q_{x_i} - q_x) n_i \psi^{(h)}_{x,i}}{\sum_i n_i \psi^{(h)}_{x,i}}$$

Using a Taylor series expansion and denoting differentials with respect to x by $'$, the BIAS term may be written as

$$\text{BIAS} = q'_x \frac{\sum_i (x_i - x) n_i \psi^{(h)}_{x,i}}{\sum_i n_i \psi^{(h)}_{x,i}} + \tfrac{1}{2} q''_x \frac{\sum_i (x_i - x)^2 n_i \psi^{(h)}_{x,i}}{\sum_i n_i \psi^{(h)}_{x,i}} + \ldots$$

$$= \text{A} + \text{B say.}$$

If h is small, $\psi^{(h)}$ is restricted so that only cases of x_i close to x contribute to the summations in terms A and B. Hence the BIAS term will be small.

The B term in the above equation will be small if q''_x is small i.e. the curve of the true population values, q_x, is approximately linear in the neighbourhood of x, or if h is small so that the effective size of $(x_i - x)^2$ in the summation is limited. The coefficient of q''_x is clearly always positive, so that if the curve of the true population values is convex, the B term will be positive and $\hat{q}_x$ will tend to overestimate q_x. The converse is true if the q_x curve is concave. This effect will be of limited importance if h is small (Figure 1).

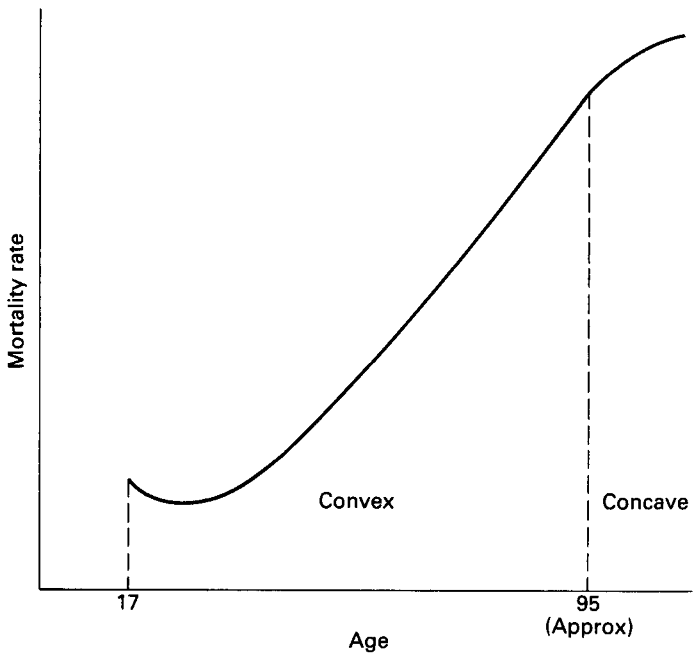

Figure 1. Male Adult Mortality Rates.

The q'_x term (A) will be zero if the values of x_i are symmetrically located about x. In practice, this term will be small at 'most' ages. However, this is not the case near the ends of the age range. For x at the lower limit all the $x_i > x$, producing positive bias at a point where the curve tends to be convex and therefore already overstated due to the bias of the q''_x term. For x at the upper limit of the age range, each $x_i < x$, the negative bias caused by this effect reinforcing that due to the fact that the curve tends to be concave here. Again the effect is reduced by decreasing the value h.

Scanty data at high and low ages tends to produce widely fluctuating crude rates $\mathring{q}_x$. The method reproduces these fluctuations in the graduated rates, rendering the ends of the curve unreliable. The full curve may simply be truncated to exclude ages with few recorded deaths. This eliminates the waves at each end of the curve but allows observations from the extremes of age to contribute to the estimated rates at more central ages.

Alternatively, the age range could be restricted before graduation. The first term of the bias expansion would then cause the graduated curve to deviate from the true population value at the new upper and lower age limits.

Another approach might be to group scanty data. The grouped values of exposed to risk and actual deaths could be attributed to the midpoint of the group, other ages from the group being excluded from the summation. The effect of this grouping of data depends on the size of h in relation to the width of the group.

When h is large, the grouped data will slightly modify the values of $\hat{q}_x$ at the end of the main block of data. The value of $\hat{q}_x$ for the central age in the group will be strongly influenced by observations towards the limit of the ungrouped data. The correlation, thus produced, between the estimated mortality rates for the grouped data and individual ages may produce an artificial plateau in the graduated curve.

If h is small in relation to the group width, the weight attached to the grouped data when estimating mortality rates towards the limit of the ungrouped data, will also be small. Thus the grouped data may contribute virtually nothing to the ungrouped estimates and, up to the end of the ungrouped data, the graduated curve may effectively be the same as that which would result from truncating the data. The ungrouped data will have little influence on the estimate for the grouped data.

3. OUTLINE OF INVESTIGATION

3.1 *The data*

Extending the work of Copas and Haberman,[1] two large data sets, already parametrically graduated, have been selected for further investigation of the kernel method. They are the standard tables of mortality for (Male) Assured Lives 1967–70[2] and those for Female Assured Lives 1975–78.[3]

Four sets of data relating to males and three sets relating to females have been graduated using kernel methods. Initially, graduation was performed over the complete age range of each parametric graduation. Where scanty data had adverse effects on the kernel graduations, the data were truncated and regraduated. This avoided the problems of grouped data outlined in section 2. Graduations have been compared from the viewpoints of goodness of fit and smoothness.

3.2 *Testing fidelity or goodness of fit*

The fidelity of kernel graduations to the original data has been assessed with the tests applied to the A1967–70 parametric graduations.[2] The standardized deviation between actual and expected deaths was calculated at each age:

$$z_x = \frac{s_x - n_x\hat{q}_x}{\sqrt{n_x\hat{q}_x(1-\hat{q}_x)}}.$$

From these deviations, a chi squared value was calculated leading to a statistic $t(\chi^2)$ which was tested at the 5% level. Since the kernel estimate requires assigning a value to only one constant, $(n-1)$ degrees of freedom were assumed, where n is the number of ages for which data were to be graduated.

To assess their randomness, the deviations were also subjected to a runs test and one for serial correlation (leading to respective test statistics $t(r)$ and $t(\rho)$). Normal approximations were used at the 5% level. The testing was one-tailed since few runs and positive correlations are undesirable, while their converses are of lesser concern and are not features of a poor fit. Further details are given in Appendix I.

An interval was identified containing values of h which produce graduations satisfying all three tests. Within this interval, the following 'rule of thumb' was used: the graduation having the smallest sum of absolute test statistics was considered to have produced the best fit.

3.3 *Testing smoothness*

In order to test for smoothness, it is necessary to decide upon the criteria or characteristics of smoothness.

The textbook by Benjamin and Pollard[4] suggests that the third differences of the graduated curve, $\Delta^3\hat{q}_x$ should be *smooth* and small. This is a circular definition.

Barnett[5] in a recent paper provides a thorough discussion of the criteria of smoothness, recognizing the cyclical nature of some of the arguments. Among his conclusions are that:

(i) second differences should pass through zero no more often than $1+2n$ times where n is the number of acceptable inherent inflections or occurrences of roughness;
(ii) a series may be regarded as smooth to the k^{th} order if k^{th} differences are insignificant, bearing in mind the decimal place at which the figures in the actual series have been truncated.

It follows from (ii) that smoothness depends partially on the scale used. We shall approach (ii) by using the relative measure

$$\left|\frac{\Delta^k \hat{q}_x}{\hat{q}_x}\right|$$

Choosing limits for this ratio to measure smoothness over different k and x is a matter of judgement. Barnett[5] suggests a target of $1/7^k$ on the grounds 'that 1 is small in relation to 7'. This is a difficult area. '1 is small in relation to 3 or 4'. Further $7^k/4^k$ and $7^k/3^k$ both $\to\infty$ as $k\to\infty$, as indicated by the figures below rounded to the nearest integer:

	k: 1	2	3	4	5
$(7/4)^k$	2	3	5	9	16
$(7/3)^k$	2	5	13	30	69

The choice of $1/7^k$ has an appealing aura of exactness which we believe is misleading. So, why $1/7^k$? Barnett[5] does not make it clear why his criterion (ii) should be based on powers of 7.

We investigate some standard, graduated mortality tables in order to establish what requirements on the smallness of differences of $\hat{q}_x$ may have been used in practice and what the results have been. In particular we consider the magnitude of $|\Delta^3\hat{q}_x|$ in each case, in absolute terms and relative to the corresponding value of $\hat{q}_x$. Table 3.3 below summarizes the results for the two life assurance mortality tables studied in this paper and for six population life tables. Values of $\max|10^5\Delta^3\hat{q}_x|$ are given, as in Barnett[5] (but note that his comments on ELT number 13 seem to be in error). In addition the value of A satisfying

$$\frac{1}{\mathrm{A}^3} = \left|\frac{10^5\Delta^3\hat{q}_x}{10^5\hat{q}_x}\right|$$

for different x and the minimum value of A are tabulated. Ages under 20 were excluded in this analysis, as were any ages for which $|10^5\Delta^3\hat{q}_x| < 2$ (to allow for rounding error). For ELT 12 ages under 27 were excluded because the parametric curve was fitted only to ages beyond this point. The parametric graduations in Table 3.3 all show small values of $\Delta^3\hat{q}_x$ and values of A in excess of 7. The A1967–70 graduation seems to be smoother than the others, which might be expected given the small number of parameters used in the graduating curve. ELT 10 and ELT 13 were graduated by non-parametric methods (using osculatory interpolation and natural cubic splines respectively) and exhibit less smoothness, in particular for ELT 10. ELT 13 is reasonably smooth beyond about age 45 for both sexes—at these higher ages the minimum values of A are 7·7 for males and 6·0 for females. It would appear that Barnett's[5] choice of powers of 7 for criterion (ii) is not unreasonable, given the actual smoothness achieved for

Table 3.3. *Measures of Smoothness*

Standard table	Type of Graduation	$\text{Max}\lvert 10^5\Delta^3 q_x\rvert$	Selected Values of A: A	Selected Values of A: x	Min A
FA1975–78	Parametric (5 par)	91 (ages 101, 102)	7·9	83	7·9
			8·3	101	
A1967–70	Parametric (4 par)	10 (ages 101–104)	10·9	66	10·9
ELT 13 Males	Splines	7 (age 96)	3·5 to 3·7	22, 27–29	3·5
			8·6	59, 62	
			9·7	70	
ELT 13 Females	Splines	12 (age 96)	3·1 to 3·3	30–32	3·1
			3·6 to 3·7	35–36	
			4·7	42	
ELT 12 Males	Parametric (7 par)	12 (age 87)	7·1	60	7·1
			8·0	66	
			12·1	87	
ELT 12 Females	Parametric (7 par)	23 (age 98)	7·7	65	7·7
			9·2	72	
			12·4	98	
ELT 10 Males	Osculatory Interpolation	161 (age 86)	4·3 to 4·4	21, 24	4·3
			5·2	86	
			5·5	85	
			6·6	58	
ELT 10 Females	Osculatory Interpolation	208 (age 86)	4·5	86	4·5
			4·7	85	
			5·1	38	
			6·3 to 6·5	53, 65, 71	

example by FA1975–78 and A1967–70. But such a choice would be likely to be too strict for a non-parametric method like the Kernel method described here.

As part of the assessment of smoothness of a kernel graduation, we calculate, therefore, second, third and fourth differences and, following criterion (i) above, second differences are checked for changes of sign. Regarding criterion (ii), the ratio $\lvert\Delta^k\hat{q}_x/\hat{q}_x\rvert$ is examined for different x and $k=2$, 3 and 4. A target of $1/\text{A}^k$ is used with A in the range 4 to 7. In the actual computation of differences, we follow Barnett's[(5)] suggestions (paragraph 6.1 of his paper) on dealing with the results of rounding error. For criterion (i), when looking at the sign changes of second differences, any values of 1 should be regarded as if they were zeros, and similarly any values of 2, providing they are not too frequent. For criterion (ii), it is best to discard one decimal place and to note that the last digit may be 'out' by 1 or 2. Again, once a stage is reached at which differences are insignificant there is no point in going any further as this would only magnify the errors present and possibly make later differences increase dramatically.

Second and third differences of $ln(\hat{q}_x)$ were investigated, the size of changes between adjacent second differences and the size of third differences being noted. Barnett[(5)] does no more than to state that these should be small. Interpreting the comments that he makes about a few graduations leads to the supposition that less than five is small. Since the differences of $ln(\hat{q}_x)$ reveal the same features as the differences of $\hat{q}_x$ with each data set, only the latter are discussed in the following sections.

3.4 *Duplicates*

The presence of duplicate policies in the ultimate data sets alters the standard deviation of the distribution of the observed deaths and hence the values of the standardized deviations, z_x. Therefore, adjustment of the χ^2 value before testing is appropriate. The method of adjustment is outlined in section 5.3, while results are summarized in sections 5.3 and 7.3.

3.5 *Two stage kernel graduations*

A modified formula for kernel graduation, incorporating prior estimates of the mortality rates, is described by Copas and Haberman.[1] This is based on the following estimating equation:

$$\hat{q}^*_x = \hat{q}^*_x + \frac{\sum_i (Z_i - q^*_{xi}) n_i \psi^{(h)}_{x,i}}{\sum_i n_i \psi^{(h)}_{x,i}}$$

where q^*_x is a prior estimate of the mortality rate at age x. This approach involves subtracting q^*_x from the data, smoothing what is left using the kernel method and then adding back q^*_x to the result to form the final estimate. Sections 8 and 9 summarize investigations into the use of this essentially two stage approach with the male and female data sets, respectively. If q^*_x is sufficiently close to the true curve, then the bias of $\hat{q}^*_x$ will be smaller than the bias of $\hat{q}_x$. Because of this reduction in bias, a larger value of h can be used, leading to a smoother graduated curve and a smaller sampling variance.[1]

3.6 *Serial correlations*

The standardized deviations calculated for kernel and parametric graduations, have been compared through the use of correlograms. Discussion of this can be found in section 9.

4. A1967–70—PARAMETRIC GRADUATION[2]

These tables were prepared using the formula:

$$\hat{q}_x = \frac{\mathrm{A} - Hx + \mathrm{B}c^x}{1 + \mathrm{A} - Hx + \mathrm{B}c^x}.$$

The data for ages over 90 appeared unreliable, so the curves were only fitted up to age $89\frac{1}{2}$ ($88\frac{1}{2}$ at duration 0). Scanty data at young ages also caused problems. All the available information was used to produce a curve which was then cut off at age 17. Below this age mortality rates were inserted based on the English Life Table No. 12 (Males).

The graduations are very smooth, none producing a second difference sign change within the age range used. All the third and fourth differences satisfy Barnett's ratio criterion viz.

$$\left|\frac{\Delta^k \hat{q}_x}{\hat{q}_x}\right| \leqslant \frac{1}{7^k}$$

as noted in section 3.3 and Table 3.3 earlier. Applying the goodness-of-fit tests to the table rates yields statistics within the 5% limits, except for the $t(X^2)$ value in the two ultimate tables. This result reflects the presence of duplicate policies and adjustments were made to the X^2 value, based on the distribution of such policies in 1954[(6)]. It was assumed the distribution of duplicates among the actual deaths was the same as that for the exposed to risk and that little or no change had taken place in the distribution between 1954 and 1967–1970. With these adjustments, $t(X^2)$ was reduced from 2·8 to ·6 for the ultimate data.

The crude data for the two standard tables considered here (A1967–70 and FA1975–78) were collected centrally from the contributing life offices by the Continuous Mortality Investigation Bureau. As has been mentioned earlier, the data are policy-based. It should also be noted that the data are heterogeneous in two senses—firstly, the data from different life offices are mixed together and secondly the classical calendar year method of investigation mixes up the experience of different cohorts (i.e. at durations over 2 years). These effects may lead to irregularities in the resulting crude rates. Hence, the crude rates may in practice be biased, contrary to the intentions outlined in section 2. Graduation is not intended to deal with the removal (or smoothing out) of such biased errors.

5. A1967–70—KERNEL GRADUATION

5.1 *Introduction*

Kernel graduations were first attempted over the same age ranges as in the parametric graduations. Estimated mortality rates were obtained for the data relating to duration 0, durations 2 and over and durations 5 and over but, at duration 1, scanty data from age $85\frac{1}{2}$ onwards produced cancelling errors. Therefore, the last five years of the data were disregarded. No attempt was made to adjust the curves below age 17.

The range of values yielding acceptably fitting curves was found for each data set. Graphs of the best fitting curves revealed distortion at the extremes of the age ranges, particularly for the select data. Therefore, a second kernel graduation was carried out using restricted age ranges.

By excluding ages with fewer than five actual deaths, the following new ranges were produced: $15\frac{1}{2}$–$65\frac{1}{2}$ at duration 0, $16\frac{1}{2}$–$70\frac{1}{2}$ at duration 1, $17\frac{1}{2}$–$89\frac{1}{2}$ for durations 2 and over and $20\frac{1}{2}$–$89\frac{1}{2}$ for durations 5 and over. For durations 2 and over this slightly reduced the interval containing values of h which produce adequately fitting graduations. For the other sets of data the interval was widened slightly.

Table 5.2a. *Test Statistics, d=0, Full Age Range.*

h	$t(\chi^2)$	$t(r)$	$t(\rho)$
1·50	−2·173	−·604	−2·244
2·00	−1·339	·338	−1·401
2·25	−·945	·327	−1·059
2·50	−·536	·327	−·735
2·75	−·099	·327	−·424
3·00	·375	·327	·121
3·25	·820	1·234	·176
3·50	1·456	1·234	·460
3·75	2·073	1·234	·742

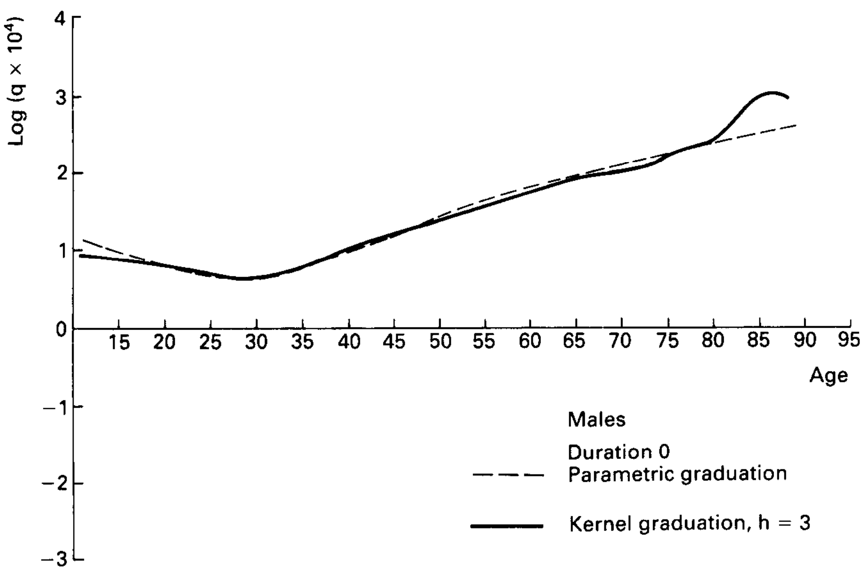

Figure 2.

5.2 *Results*

At duration 0, $h=3$ produces the best fitting curve for the full data (Table 5.2a and Figure 2) and also the truncated data (Table 5.2b and Figure 3). Investigating Barnett's ratio criterion with A=7, the full curve is found to be smooth for the first half of the age range but poorly smoothed at high ages. The truncated curve is not well smoothed even at young ages and it is necessary to

Table 5.2b. *Test Statistics, d=0, Restricted Age Range*

h	$t(\chi^2)$	$t(r)$	$t(\rho)$
1·00	−2·550	−1·437	−3·269
1·50	−1·579	−1·356	−2·371
2·00	−1·022	−·670	−1·623
2·50	−·448	−·654	−·873
2·75	−·103	−·654	−·483
3·00	−·293	−·654	−·085
3·25	·745	·056	·314
3·50	1·257	·056	·707
4·00	2·472	·056	1·457

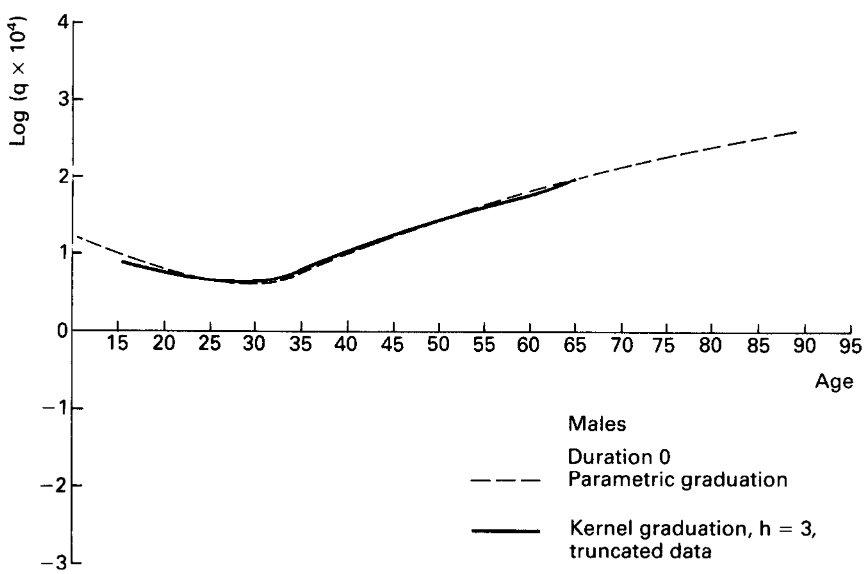

Figure 3.

increase h to 8 before all third and fourth differences satisfy the inequality. Setting A to 4 produces values of the ratio which indicate that the full curve is smooth until the mid sixties of age. The truncated curve now appears to be perfectly smooth.

Cancelling errors resulting from scanty data at high ages force some restriction on the age range of the original graduation for duration 1 (Table 5.2c). Further restriction is imposed by the criterion of at least five deaths at each age in the second graduation (Table 5.2d). In each case the best curve results from $h=2{\cdot}75$ (Figures 4 and 5).

Table 5.2c. *Test Statistics, $d=1$, Age Range $10\frac{1}{2}$–$84\frac{1}{2}$*

h	$t(\chi^2)$	$t(r)$	$t(\rho)$
1·00	−2·574	−2·909	−4·143
1·50	−1·302	−2·459	−2·829
2·00	−·390	−2·071	−1·820
2·50	·479	−·786	−1·030
2·75	·942	−·390	−·670
3·00	1·447	−·390	−·312
3·25	2·009	·000	·036

Table 5.2d. *Test Statistics, $d=1$, Restricted Age Range*

h	$t(\chi^2)$	$t(r)$	$t(\rho)$
1·00	−2·606	−3·219	−4·025
1·50	−1·581	−2·585	−2·645
2·00	−·797	−2·724	−1·514
2·50	−·018	−1·339	−·615
2·75	·398	−1·339	−·213
3·00	·847	−·943	·175
3·50	1·895	·263	·934

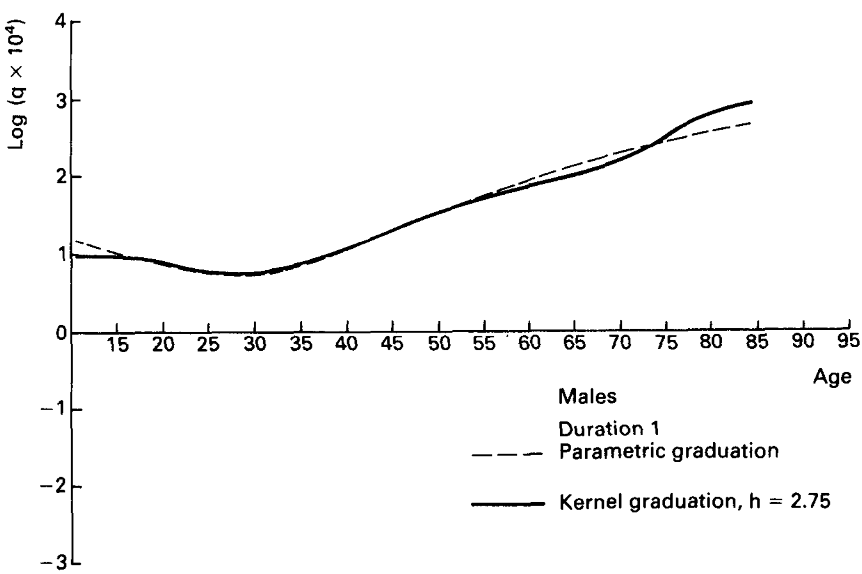

Figure 4.

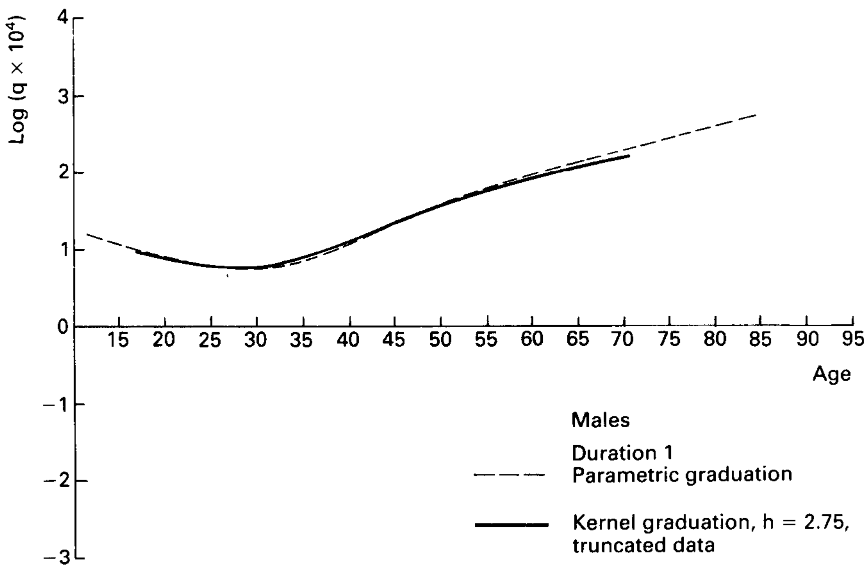

Figure 5.

These curves are smoother than their counterparts at duration 0. The longer one has three second difference sign changes and the majority of the third differences until age $64\frac{1}{2}$ satisfy the ratio test with $A=7$. However, most of the fourth differences fail the test. The differences for the truncated curve follow the same general pattern but those, satisfying the ratio test, are split into a larger number of groups. Hence, the truncated curve is slightly less smooth than the fuller one. Third order smoothness is obtained in the longer curve by increasing h to 6.

Repeating the investigation of smoothness with $A=4$ yields ratio results which suggest that the full curve is smooth up to age 70. The truncated curve is smooth, only producing one fourth difference which fails to satisfy the inequality in Barnett's ratio criterion.

The best curves for the full and restricted age ranges at durations 2 and over, both result from $h=1{\cdot}2$ (Tables 5.2e and 5.2f and Figures 6 and 7). If $A=7$ is used in Barnett's smoothness criterion, these curves appear to be very poorly smoothed. Only by increasing h to 7 is third order smoothness obtained. When $A=4$ is inserted in the ratio, the latter half of the third differences for both curves, satisfy the inequality. However, the other differences indicate poor smoothing.

Very poor smoothing is again a feature of the best fitting curves for durations 5 and over. Very few 3rd or 4th differences satisfy Barnett's ratio criterion when A equals 7. Reducing A to 4 still leaves half of these differences failing the test. The

Table 5.2e. *Test Statistics, d⩾2, Full Age Range.*

h	$t(\chi^2)$	$t(r)$	$t(\rho)$
0·8	−2·469	−3·158	−3·981
0·9	−1·596	−3·158	−3·391
1·0	−1·107	−1·828	−2·698
1·1	·020	−·925	−1·936
1·2	·859	−·248	−1·132
1·3	1·773	−·248	−·309

Table 5.2f. *Test Statistics, d⩾2, Restricted Age Range.*

h	$t(\chi^2)$	$t(r)$	$t(\rho)$
0·8	−2·050	−3·712	−3·957
0·9	−1·146	−3·712	−3·396
1·0	−·304	−2·407	−2·739
1·1	·528	−1·450	−2·015
1·2	1·398	−1·286	−1·247
1·3	2·347	−1·286	−·457

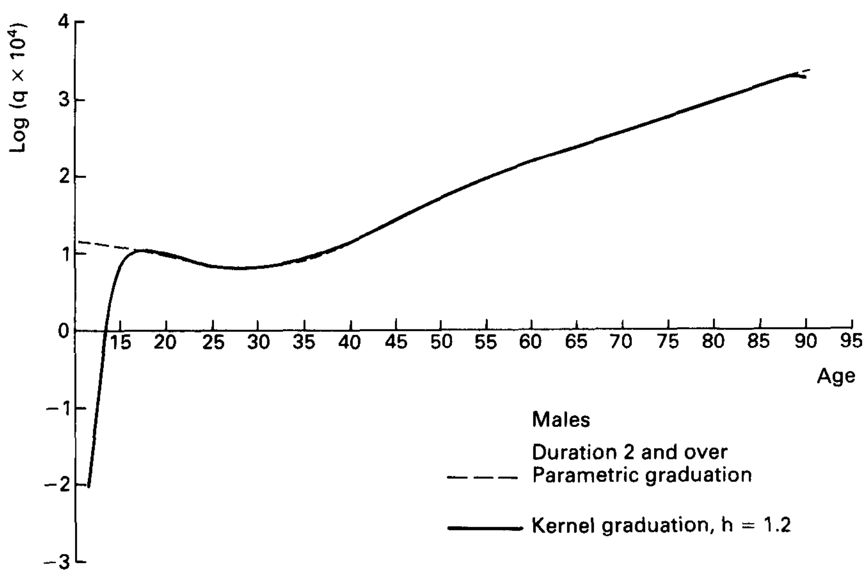

Figure 6.

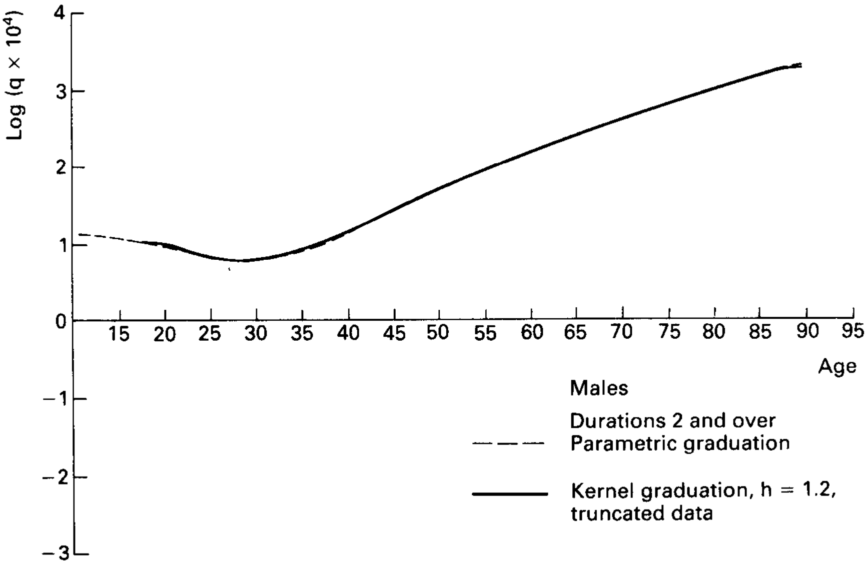

Figure 7.

best full curve results from $h = 1{\cdot}2$, (Figure 8) while the best truncated curve is $h = 1{\cdot}1$ (Figure 9 and Tables 5·2g and 5·2h). Third order smoothness when A = 7 is obtained by increasing h to 7.

5.3 *Adjustment for duplicates*

The parametric graduations for the A1967–70 data were adjusted to compensate for the presence of duplicate policies in the ultimate data[6] using variance ratios r_x. The kernel graduations of the ultimate data have been treated in the same way. New standardized deviations were calculated using $z'_x = z_x/\sqrt{r_x}$. Then the adjusted chi squared value, $\chi'^2 = \sum (z'_x)^2$ was used to test the fidelity of the graduated rates. Two series of ratios are given[6], one for all offices contributing to the investigation and one for non-industrial offices reporting 300 or more policy claims. The latter was used with the kernel graduations.

The adjustment results in larger values of h producing adequately fitting curves (Tables 5.3a and 5.3b). For durations 2 and over the best fitting curve uses $h = 1{\cdot}3$ (Figure 10) rather than $h = 1{\cdot}2$ without adjustment. There is a slightly larger increase in the optimum value of h for durations 5 and over, from 1·1 to 1·4 (Figure 11). Both of the new best fitting curves are very poorly smoothed, judged against the A = 7 criterion.

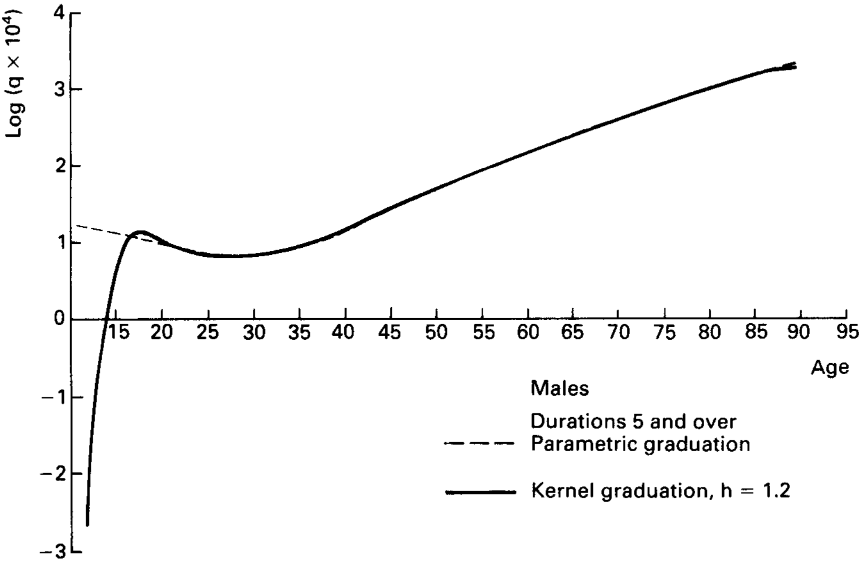

Figure 8.

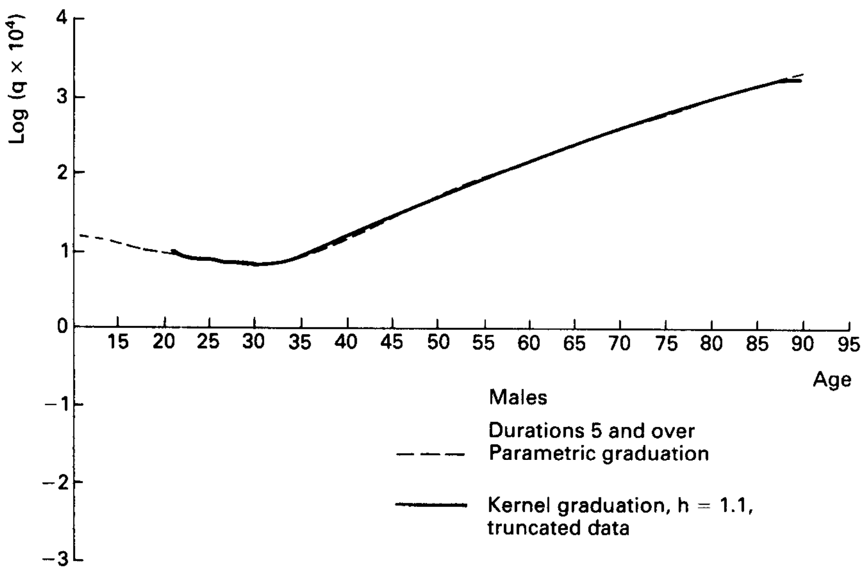

Figure 9.

Table 5.2g. *Test Statistics, d⩾5, Full Age Range.*

h	$t(\chi^2)$	$t(r)$	$t(\rho)$
0·8	−2·357	−3·274	−3·884
0·9	−1·454	−3·274	−3·314
1·0	−·868	−2·510	−2·650
1·1	·298	−2·142	−1·924
1·2	1·062	−1·678	−1·159
1·3	1·979	−·143	−·375

Table 5.2h. *Test Statistics, d⩾5, Restricted Age Range.*

h	$t(\chi^2)$	$t(r)$	$t(\rho)$
0·7	−2·864	−3·772	−4·070
0·8	−1·812	−3·772	−3·661
0·9	−·863	−3·772	−3·124
1·0	·024	−3·112	−2·498
1·1	·895	−2·753	−1·811
1·2	1·796	−2·404	−1·058

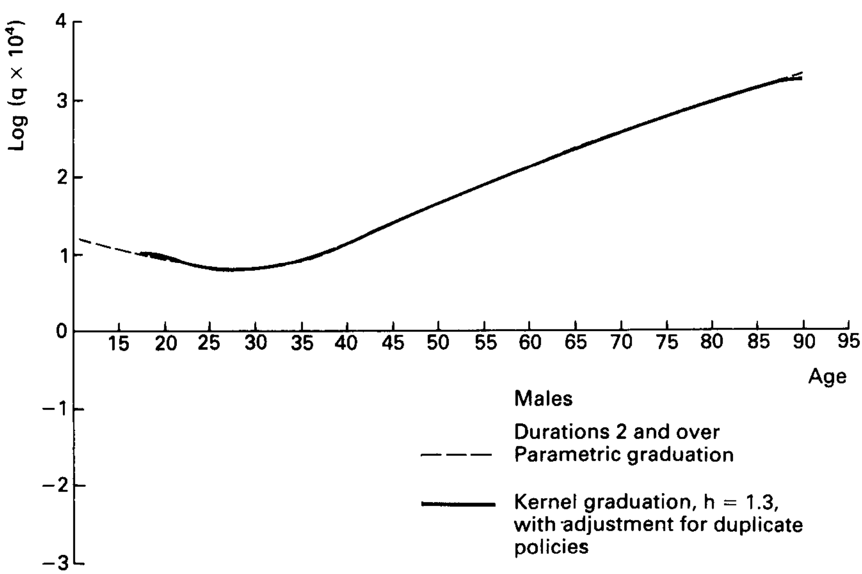

Figure 10.

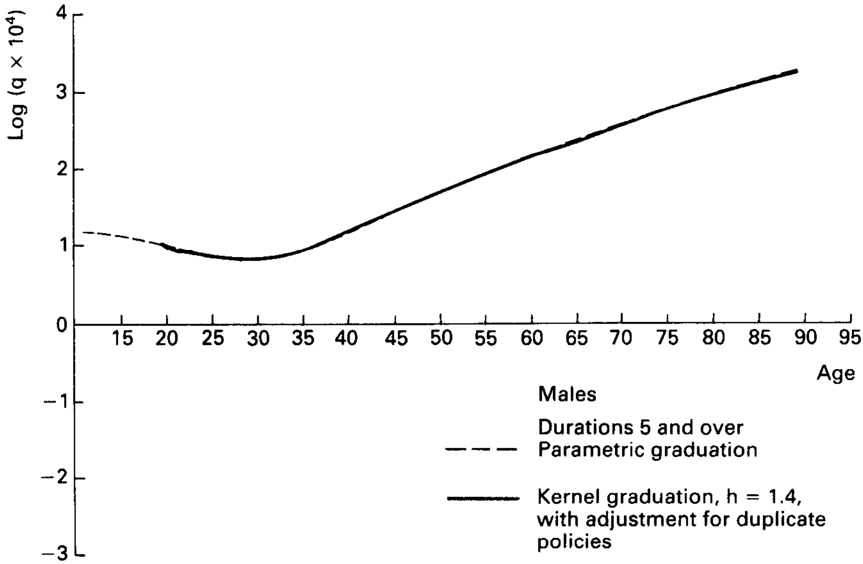

Figure 11.

Table 5.3a. *Test Statistics, $d \geqslant 2$, With Duplicates Adjustment*

h	$t(\chi'^2)$	$t(r)$	$t(\rho)$
1·1	−2·070	−·925	−1·936
1·2	−1·496	−·248	−1·132
1·3	−·869	−·248	−·309
1·4	−·165	1·081	·513
1·5	·640	1·081	1·312
1·6	1·565	2·008	2·074

Table 5.3b. *Test Statistics, $d \geqslant 5$, With Duplicates Adjustment*

h	$t(\chi'^2)$	$t(r)$	$t(\rho)$
1·1	−1·907	−2·142	−1·924
1·2	−1·322	−1·678	−1·159
1·3	−·692	−·143	−·375
1·4	·007	·175	·409
1·5	·796	·543	1·177
1·6	1·694	1·533	1·915

6. FA1975–78—PARAMETRIC GRADUATION[3]

The ultimate rates in these tables were obtained from the formula,

$\hat{q}_x = e^{pol(x)}/1 + e^{pol(x)}$
where
$pol(x) = a_1 + a_2 t + a_3(2t^2 - 1) + a_4(4t^2 - 3t) + a_5(8t^4 - 8t^2 + 1)$ and $t = x - 70/50$.

The select rates were obtained from the ultimate rates by deducting 6·51725591 years at duration 0 and 5·07221888 years at duration 1 from the respective age. The ultimate rates are tabulated for ages 20 to 94 while select rates are given for ages from 20 to 74.

The differences of the table mortality rates indicate very smooth curves. There are no sign changes in the second differences and all third and fourth differences satisfy the ratio test (for A = 7). Testing the fit of these graduations produces satisfactory test statistics, except for $t(\chi^2)$ at durations 2 and over. (i.e. $t(\chi^2) = 2{\cdot}25$ for the ultimate data). Adjusting this statistic for the presence of duplicates is not discussed in the C.M.I. report.[3]

7. FA1975–78—KERNEL GRADUATION

7.1 *Introduction*

Initially, each set of data was used in full to produce kernel graduations. This resulted in acceptable curves from the ultimate data, but badly distorted curves from the select data. Therefore, the latter were regraduated after discarding the scanty data at high ages.

The age range chosen at duration 0 was 20–56, which includes eight ages at which the number of deaths is fewer than five. These all lie in the first half of the age range. At duration 1 the age range 20–61 was used, including six ages with fewer than five deaths, all early in the table. For each set of data, the kernel mortality rates lie above the table rates at young ages and drop below towards the end of the table. This is referred to in section 7.4.

7.2 *Results*

At duration 0 the second group of kernel graduations (on the restricted age range) spanned only 37 ages. This introduced the possibility of both n_1 and n_2 falling below 20, rendering the normal approximation for the runs test unreliable. However, this occurred in only two graduations, $h = 1$ and $h = 2{\cdot}5$, (Table 7.2a).

Ignoring the runs test, we find that the best fitting graduation results from $h = 4$ (Figure 12). If the runs test is included, $h = 2{\cdot}5$ produces the best graduation. Setting A equal to 7, the curve using $h = 2{\cdot}5$ shows very poor smoothing with six sign changes in the second differences and very few third or fourth differences satisfying the ratio test. The $h = 4$ curve is much smoother, having just two second difference sign changes, with all third differences and the majority of the fourth differences satisfying the ratio test.

Increasing h to 5, which still yields a well-fitting graduation, produces a

Table 7.2a. *Test Statistics, d=0, Restricted Age Range*

h	$t(\chi^2)$	$t(r)$	$t(\rho)$
·5	−3·976	−2·559*	−3·078
1·0	−1·846	−·839	−2·450
2·0	−·169	−·209	−1·123
2·5	·173	·496*	−·787
3·0	·415	1·134	−·536
3·5	·613	1·134	−·324
4·0	·803	1·134	−·121
4·5	1·006	1·134	·088
5·0	1·230	·735	·310
6·0	1·767	·735	·797

* Both n_1 and n_2 below 20.

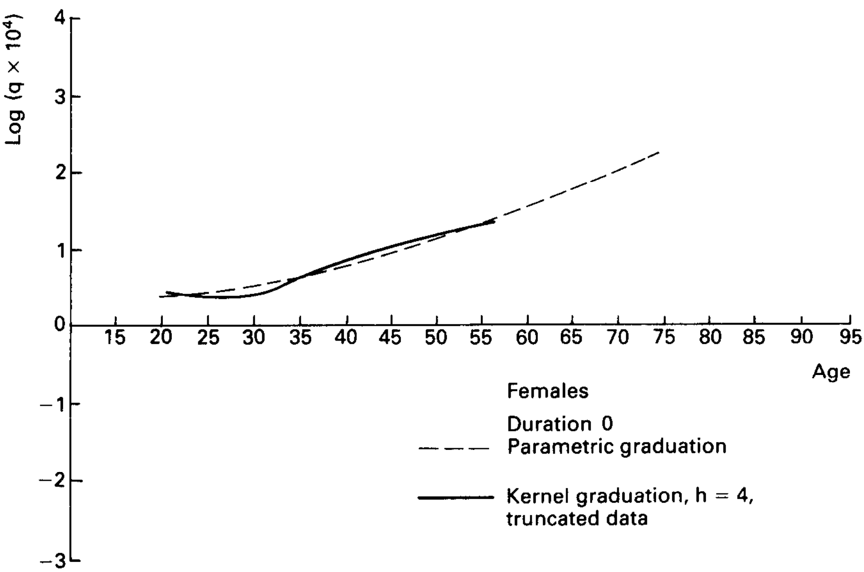

Figure 12.

Table 7.2b. *Test Statistics, d=1, Restricted Age Range*

h	$t(\chi^2)$	$t(r)$	$t(\rho)$
2·5	−2·139	−1·870	−1·121
3·0	−1·916	−1·390	−·725
4·0	−1·443	−1·870	−·046
4·5	−1·174	−1·870	·281
5·0	−·876	·077	·608
5·5	−·543	·821	·935
6·0	−·171	·821	1·260
6·5	·242	2·097	1·579

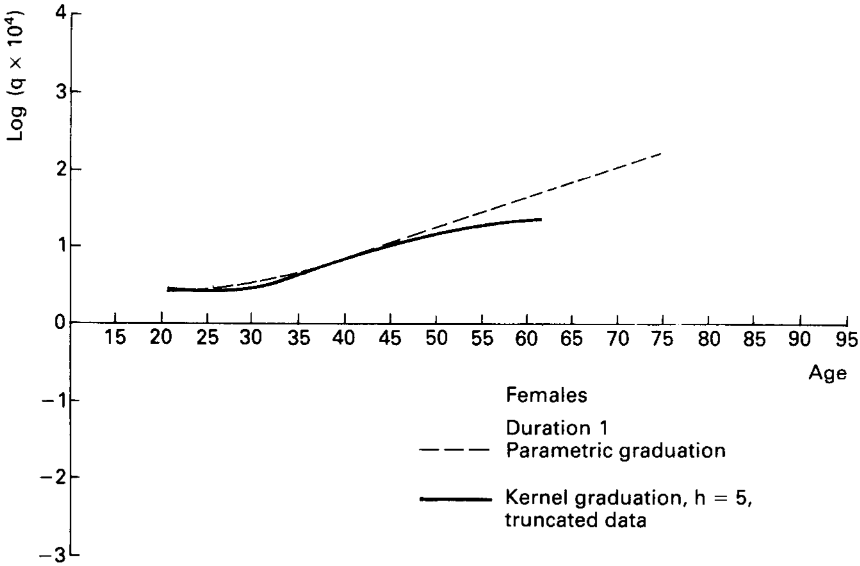

Figure 13.

smooth curve. There is one second difference sign change and all third and fourth differences satisfy the ratio test. Moving outside the goodness-of-fit limits to $h=7$ leads to a graduation with zero third and fourth differences.

If A=7 is replaced by A=4, the $h=2{\cdot}5$ curve still appears to be poorly smoothed. However, the $h=4$ curve is now perfectly smooth.

The restricted age range chosen at duration 1 contained 42 ages, so there were no problems with the runs test. The best fitting curve uses $h=5$ (Table 7.2b and Figure 13). It is smooth, having one second difference sign change, third differences which satisfy the ratio test (based on A=7) throughout and zero fourth differences.

Table 7.2c. *Test Statistics, $d \geqslant 2$, Full Age Range*

h	$t(\chi^2)$	$t(r)$	$t(\rho)$
1·0	−2·158	−3·893	−5·526
1·5	−1·067	−3·391	−4·394
2·0	−·355	−2·774	−3·340
2·5	·484	−1·331	−2·237
3·0	1·654	−1·470	−1·009

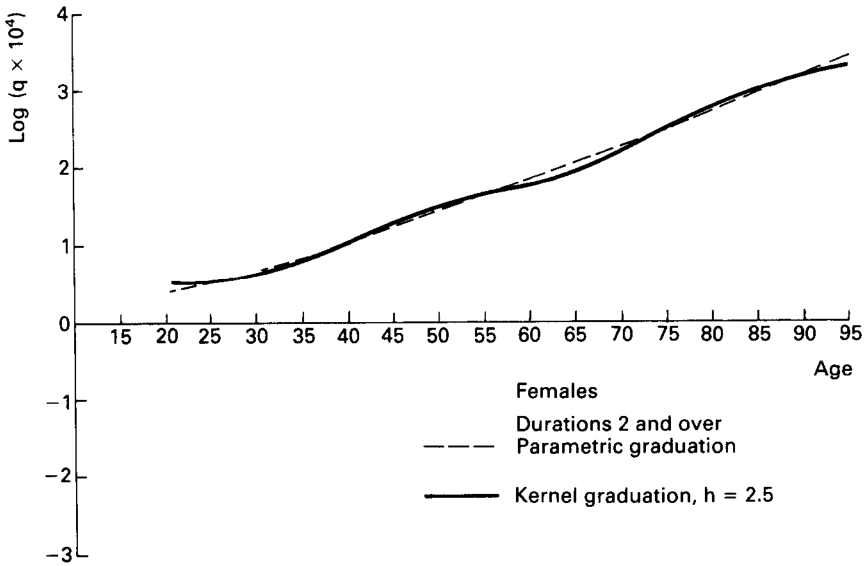

Figure 14.

For durations 2 and over there are more than five deaths at each age; therefore the kernel graduations used all the data available. The best curve results from $h=2{\cdot}5$ (Table 7.2c and Figure 14). The smoothness criterion suggests that this curve is reasonably smooth, having one second difference sign change late in life. The majority of the third and fourth differences satisfy the ratio test. However, h must be increased to 9 before the ratio test is satisfied by all third and fourth differences, there remaining one sign change in the second differences. If smoothness is tested using $A=4$, the curve has third order smoothness and the majority of the fourth differences satisfy the appropriate inequality.

7.3 *Adjustment for duplicates*

The ultimate data were regraduated using the variance ratios described in section 5.3, to adjust the χ^2 value. This slightly altered the interval containing values of h which produce well-fitting graduations (Table 7.3). The best curve now results from $h=3{\cdot}5$ (Figure 15). It is smoother than the unadjusted curve, but there remain some third and fourth differences which do not satisfy the smoothness inequality with $A=7$.

Figures 12–15 indicate that the best kernel graduations have waves, although they are smooth. These waves correspond to irregularities in the crude data (mentioned briefly in section 5) and may be real features exhibited by the underlying population mortality rates. The parametric graduations take no

Table 7.3. *Test Statistics, d=0, With Duplicate Policies Adjustment*

h	$t(\chi'^2)$	$t(r)$	$t(\rho)$
2·0	−2·048	−2·774	−3·342
2·5	−1·462	−1·331	−2·243
3·0	−·651	−1·469	−1·018
3·5	·501	·171	·291
4·0	2·101	·171	1·585

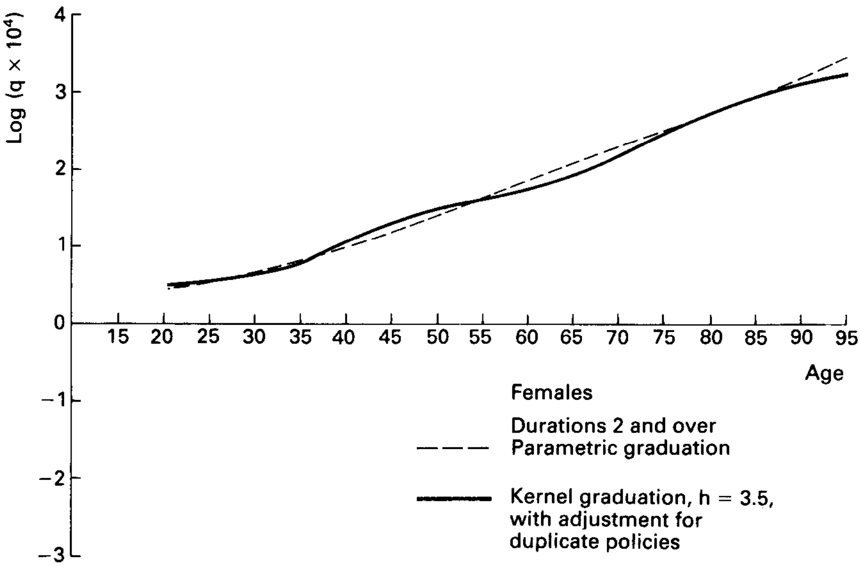

Figure 15.

account of these waves, of course, due to the mathematical properties of the functions selected.

7.4 *The youngest ages*

As has been noted earlier, at the youngest ages (i.e. under 30) the best fitting kernel graduation has rates which are greater than the table graduated rates.

For durations 0, 1 and 2 and over, the table graduated mortality rates increase monotonically from age 20 onwards. These parametric curves, fitted by the method of maximum likelihood or minimum χ^2, are influenced little by these ages where the data are scanty and in particular the numbers of deaths are under 10.

Table 7.4. *Comparison of Crude and Graduated Mortality Rates at the Youngest Ages.*

Age	$q_x \times 10^4$		
Duration 0	Crude	Parametric	Best Kernel ($h=4$)
$20\frac{1}{2}$	2·625	2·378	2·656
$21\frac{1}{2}$	2·503	2·401	2·624
$22\frac{1}{2}$	4·602	2·437	2·589
$23\frac{1}{2}$	·849	2·488	2·553
$24\frac{1}{2}$	2·634	2·552	2·521
$25\frac{1}{2}$	3·571	2·631	2·496
$26\frac{1}{2}$	1·853	2·725	2·487
$27\frac{1}{2}$	·464	2·835	2·501
$28\frac{1}{2}$	3·184	2·961	2·546
Duration 1	Crude	Parametric	Best Kernel ($h=5$)
$20\frac{1}{2}$	4·995	2·415	3·252
$21\frac{1}{2}$	5·153	2·458	3·176
$22\frac{1}{2}$	3·877	2·515	3·110
$23\frac{1}{2}$	1·432	2·585	3·057
$24\frac{1}{2}$	·971	2·671	3·018
$25\frac{1}{2}$	2·995	2·772	2·999
$26\frac{1}{2}$	3·515	2·889	3·001
$27\frac{1}{2}$	2·046	3·023	3·029
$28\frac{1}{2}$	2·569	3·176	3·087
$29\frac{1}{2}$	3·128	3·349	3·178
Durations 2 & over	Crude	Parametric	Best Kernel ($h=2·5$)
$20\frac{1}{2}$	3·418	2·780	3·258
$21\frac{1}{2}$	4·381	2·898	3·220
$22\frac{1}{2}$	2·666	3·033	3·232
$23\frac{1}{2}$	2·063	3·188	3·306
$24\frac{1}{2}$	2·856	3·362	3·438

Indeed, the crude rates fluctuate markedly at these ages (see Table 7.4). In smoothing out these fluctuations, the kernel method leads to sets of graduated mortality rates that decrease with increasing age and so have a different shape from the FA1975–78 curves but a *similar* shape to the corresponding male A1967–70 curves. Table 7.4 indicates that the minimum kernel mortality rates are achieved at ages: $26\frac{1}{2}$ for $d=0$; $25\frac{1}{2}$ for $d=1$; $21\frac{1}{2}$ for $d\geqslant 2$.

Given the fluctuations in the crude data, it is not possible to say whether this feature produced in the kernel graduations is a real feature. To quote Mr. R. Barley: "When is a wave not a wave?"

8. TWO-STAGE ESTIMATES

8.1 *Introduction*

A modified formula for calculating kernel estimates is described by Copas and

Haberman[1]. It utilizes a prior estimate of the mortality rate, q_x^*.

$$\hat{q}_x^* = q_x^* + \frac{\sum_i (Z_i - q_{xi}^*) n_i \psi_{x,i}^{(h)}}{\sum_i n_i \psi_{x,i}^{(h)}}.$$

If h is small, as noted by Copas and Haberman[1], there is little difference between $\hat{q}_x$ and $\hat{q}_x^*$. If the prior estimate, q_x^*, is sufficiently close to the true mortality rate, the bias of this modified estimate, $\hat{q}_x^*$, is smaller than that of the one-stage estimate. This permits the use of larger values of h, reducing the sampling variance of $\hat{q}_x^*$ and improving the smoothness of the graduated curve. However, finding suitable prior estimates has proved difficult.

8.2 *Two-stage estimates—A1967–70*

The first CMI report compared actual deaths in the 1967–70 data with deaths expected by the A1949–52 tables. For the select data, over all ages, the links between the two data sets were found to be:

$$d=0 \qquad \frac{\text{Actual (A67–70)}}{\text{Expected (A49–52)}} \times 100 = 86$$

$$d=1 \qquad \frac{\text{Actual (A67–70)}}{\text{Expected (A49–52)}} \times 100 = 82$$

Therefore, prior estimates of $q_x^* = \cdot 86\, q_x$ (1949–52) and $q_x^* = \cdot 82\, q_x$ (1949–52) were tried with the modified formula. For the ultimate data a more complex relationship was suggested, $q_x^* = \cdot 84\, q_x$ (1949–52) $- \cdot 00016$. This has been used as the prior estimate for durations 2 and over.

All the graduations using these prior estimates are poor. For the select data, $h = 1$ produces very large χ^2 statistics, as does $h = \cdot 5$ for the ultimate data. Using $h = \cdot 6$ produces an adequately fitting graduation for each set of select data but the graduated curves contain many small waves and so are unacceptable.

Other mortality rates which suggested themselves as possible candidates for prior estimates were those from the best one-stage kernel graduations. All four sets (durations 0, 1, 2 and over, 5 and over) give poor results, requiring smaller values of h to obtain an acceptable fit than had been used in the one-stage process. In each case the level of smoothness is reduced by the application of the two-stage method (as would be expected with smaller values of h).

The best one-stage graduations including adjustment for duplicate policies, are slightly better prior estimates than the unadjusted rates, producing fewer waves. However, for both sets of ultimate data, setting h as low as $\cdot 5$ still produces test statistics well in excess of the 5% limits.

8.3 *Two-stage estimates—FA1975–78*

Attempts were made to find suitable prior estimates for the female data. These were no more successful than those described in section 8.2. The fifth CMI report[9] has made a comparison between actual deaths in FA1975–78 data and the deaths expected by three other tables. Mortality rates from two of them have been used as prior estimates.

Firstly, the A1967–70 rates were used with a four year age deduction, since the CMI Report's authors[9] believed that the deduction represented the then current practice. Over all ages, the percentages from the comparison between actual and expected deaths were 68% at duration 0, 65% at duration 1 and 81% for the ultimate data. Prior estimates of $q_x^* = \cdot 66\, q_{x-4}$ (1967–70) for the two sets of select data and $q_x^* = \cdot 81\, q_{x-4}$ (1967–70) for the ultimate data were investigated. For each set of data, the two-stage method requires smaller values of h than the one-stage method and the newly graduated curves contain many small waves.

Percentages comparing the A1967–70 tables with the FA1975–78 data were also given for five year age groups.[9] Prior estimates utilizing these percentages were tried. The sharp change in weights between age groups appeared as a series of steps in the graduated rates. To remedy this, a smooth curve was drawn by age through the midpoints of the upper edges of the histogram bars representing the weights. Care was taken to preserve the area under the graph. A weight w_x was read from the curve for each age. Thus, the following prior estimate was used: $q_x^* = w_x\, q_{x-4}$ (1967–70).

These prior estimates were slightly more successful than using a single multiplying factor throughout the age range. However, it was still necessary to use a smaller value of h than the one-stage method required, in order to obtain an acceptable fit. Waves occurred in the graduated curves, but were not as numerous as previously.

It was noted in the CMI report[9] that the A1967–70 curve had a different shape from the female experience. The prior estimates based on this curve are not sufficiently close to the true mortality rates for the two-stage method to give improved results over the one-stage method.

The report also compared the female data for 1975–78 with the ELT 13 (Females), concluding that the tabulated population mortality rates are rather high. This indicates that the ELT 13 (Females) rates are unlikely to be good prior estimates and a few trial graduations have been carried out to verify this. The estimates were found from the rates tabulated at integral ages by interpolation. Very small values of h are required to provide an adequate fit and therefore the smoothness of the curves is unsatisfactory.

It was hoped that the one-stage kernel estimates would be better prior estimates than those based on other tables. This proved to be the case, particularly with the select data. However, the curves resulting from the two-stage method are less smooth than the coresponding one-stage kernel graduations, obviating the extra work.

The one-stage kernel estimates after adjustment for duplicate policies yield slightly improved results over those described above. However, it is still necessary to set h below $\cdot 5$, resulting in poor smoothing. Thus, the original formula is to be preferred to the two-stage method.

8.4 *Comments*

It would appear that this two-stage kernel method introduced by Copas and Haberman[1] does not work satisfactorily unless the prior estimate is very close to the true rates.

In the cases discussed in sections 8.2 and 8.3, none of the prior estimates are sufficiently close to the true rates, so that a larger value of h (than in the one-stage method) does not emerge. When h is small, there is little to choose between the one-stage and two-stage approaches and, of course, a lack of smoothness may be apparent.

Further, with large data sets like those underlying the two standard tables discussed here, A1967–70 and FA1975–78, the level of sampling variation in the crude $\mathring{q}_x$ is very small at most ages and this is an additional factor explaining why an approximate prior estimate leads to a poorly graduated curve from this method.

9. CORRELOGRAMS

Standardized deviations between actual and expected deaths, described in section 3, were calculated for each kernel graduation. If the recorded data are a random sample from a population with mortality rates equal to those given by the graduation the z's will be randomly drawn from the standard Normal distribution. This has been checked by the three tests of fit listed in section 3, including an investigation of correlation between successive z's. Graduations, not producing statistics within the 95% confidence intervals, have been rejected as not providing an adequate fit to the data.

Deviations from the results predicted by the graduated rates are equivalent to the superimposed errors referred to by Elphinstone[10]. He proposed an investigation of their randomness using correlograms and periodograms. If serial correlations for the z_x are calculated at different lags k, then provided the z_x are random, the correlations r_k will be approximately $N(0,1/n)$, where n is the number of standardized deviations.[11]

Table 9. Serial correlations falling outside the confidence intervals.

	Number of r_k's outside 95% limits	
	Parametric Graduation	Kernel Graduation
A1967–70		
$d=0$	1	0
$d=1$	0	0
$d\geqslant 2$	1	1
$d\geqslant 5$	0	1
FA1975–78		
$d=0$	0	0
$d=1$	0	1
$d\geqslant 2$	3	2

Correlograms were produced for the standardized deviations of the parametric and best one-stage kernel graduations, for each set of data. Correlations were calculated for values of k up to $(n/4)$ or $k = 15$, whichever was smaller, where n is the number of data points. This criterion has been suggested by Chatfield[11]. The 95% confidence intervals for the means were added to the graphs and Table 9 summarizes the results. The correlograms for the corresponding parametric and kernel graduations are similar in shape.

10. CONCLUSIONS

Earlier work on kernel graduation[1] has been extended by investigating its application to larger data sets. The kernel method provides an easily applied method of graduation which, through the adjustment of the constant h, can be constrained (in a controlled way) to fit the data closely, provide a high level of smoothing, or a balance between these two characteristics. The method does not involve any a priori view of the true form of the underlying rates.

The method is suitable for computer application with little prior preparation of the data. If the presence of scanty data forces truncation, the ends of the curve must be obtained by alternative means. This is a feature of other graduation methods and therefore, not a great disadvantage. However, parametric methods do allow mortality rates to be extended beyond the age range of the data.

A range of kernel graduations has been produced for seven data sets. Those satisfying the tests of fit applied to the parametric graduations are very close to the published graduations, but have been obtained more simply. On smoothness, the three female graduations and the two male select graduations satisfy the $A = 4$ criterion while only the female graduations satisfy the $A = 7$ criterion. The application of the modified formula has proved somewhat disappointing, with difficulties being encountered in identifying good prior estimates. It is possible that the modified approach is only practicable when the underlying numbers of deaths and exposed to risk are not too large and prior estimates, close to the true rates, are available.

The results are encouraging and indicate that kernel methods may be especially useful in a number of practical circumstances. Thus, the methods seem suitable when the underlying data are not too large or when an idea of the shape of the underlying population rates is required as a preliminary to a full parametric investigation. The methods have particular usefulness when the mathematical formula, describing the shape of the underlying population rates, is not known, irregular or highly complex. It should be noted that the reliance on only one adjustment constant and the data responsiveness of the methods mean that it is only to be expected that certain graduated curves are poor from either the viewpoint of smoothness or goodness-of-fit.

Parametric formula-based graduation methods are suitable in other circumstances, particularly when the underlying data are extensive. Such methods may be generalized for application to bivariate situations where there are two independent variables, for example, age- and duration-specific recovery rates for sickness or disability insurance.

11. ACKNOWLEDGEMENTS

Some of the material in this paper was presented to a meeting of the Merseyside Section of the Royal Statistical Society on 17 February 1987 and to a joint meeting of the Staple Inn Actuarial Society and the General Applications Section of the Royal Statistical Society on 3 March 1987.

REFERENCES

(1) Copas, J. B. & Haberman, S. (1983). Non-Parametric Graduation Using Kernel Methods. J.I.A. **110.** 135.
(2) Joint Mortality Investigation Committee (1974). Considerations Affecting the Preparation of Standard Tables of Mortality. J.I.A. **101,** 133.
(3) CMI Report No. 6 (1983). Graduation of the Mortality Experience of Female Assured Lives 1975–78.
(4) Benjamin, B. & Pollard J. H. (1980). The Analysis of Mortality and Other Actuarial Statistics. Chapter 11, page 242. Heinemann, London.
(5) Barnett, H. A. R. (1985). Criteria of Smoothness. J.I.A. **112,** 331.
(6) Joint Mortality Investigation Committee (1957). Continuous Investigation into the Mortality of Assured Lives: Memorandum on a Special Inquiry into the Distribution of Duplicate Policies. *J.I.A.* **83,** 34.
(7) CMI Report No. 1 (1973). Mortality of Assured Lives 1967–70.
(8) Joint Mortality Investigation Committee (1956). Mortality of Assured Lives. J.I.A. **82,** 3.
(9) CMI Report No. 5 (1981). Mortality of Assured Lives. 1975–78.
(10) Elphinstone, M. D. W. (1950). Summation and Some Other Methods of Graduation. The foundations of theory. T.F.A. **20,** 25.
(11) Chatfield, C. (1980). The Analysis of Time Series. Chapman and Hall.

APPENDIX I: TESTS OF GRADUATION

$$z_x = \frac{s_n - n_x\hat{q}_x}{\sqrt{n_x\hat{q}_x(1-\hat{q}_x)}}$$

To test whether the z's are Normally distributed, $\chi^2 = \sum z_x{}^2$ is calculated. Where the number of degrees of freedom exceeds 50, the statistic $t(\chi^2) = \sqrt{2\chi^2} - \sqrt{2(n-1)}$ is approximately Normally distributed with zero mean and unit variance. The χ^2 test assumes independence of the events at each age and an allowance needs to be made for the presence of duplicate policies—see text.

To test whether the z's are randomly arranged, two methods are possible:

(i) The Wald-Wolfowitz test was used on the number of 'runs' formed by the signs of the deviations. Assuming that there are n_1 positive and n_2 negative durations, then r, the number of runs (where a run is a sequence of deviations over successive ages all with the same sign), is approximately Normally distributed with mean μ_r and variance $\sigma_r{}^2$ where

$$\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1$$

$$\sigma_r{}^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}\ .$$

The number of runs should not be too small. Too large a number of runs would indicate that the data were peculiar, not that the fitted curve was unsatisfactory.

(ii) A serial correlation test was used on the values of the z_x's. If the z's are randomly distributed then the correlation coefficient between successive values of z (z_y and z_{y-1} for all y), ρ, is approximately Normally distributed with zero mean and variance $\frac{1}{\text{n}}$ where n is the number of age groups.

Here $$\rho = \frac{\sum_{y=x_0}^{x_0+n-2}(z_y-\bar{z})(z_{y+1}-\bar{z})}{\sum_{y=x_0}^{x_0+\text{n}-1}(z_y-\bar{z})^2} \quad \text{where} \quad \bar{z} = \frac{\sum_{y=x_0}^{x_0+\text{n}-1} z_y}{\text{n}}$$

for an age range of $(x_0, x_0+\text{n}-1)$.

Too large a positive value of ρ would indicate an unsatisfactory fit. Too large a negative value would suggest that the z's were alternatively positive and negative to too great an extent.