ON THE DUALITY OF ASSUMPTIONS UNDERPINNING THE CONSTRUCTION OF LIFE TABLES

A. E. RENSHAW, PH D , S. HABERMAN, PH.D., F.I.A., AND P. HATZOPOULOS, M.SC.
of City University of London

ABSTRACT

We investigate the implications of a dual approach to the graduation of the force of mortality based on the modelling of the exposures as gamma random variables, as opposed to the modelling of the numbers of deaths as Poisson random variables.

KEYWORDS

Graduation, Life Tables; Exposure Response Models; Generalised Linear Models

1 INTRODUCTION

In this paper, we describe as the 'conventional' approach to graduation the method whereby the force of mortality is graduated by fitting a parameterised formula to the crude mortality rates under the assumption that the actual numbers of deaths are Poisson random variables conditional on the matching central exposures to the risk of death, e.g. Forfar, McCutcheon & Wilkie (1988) Under this approach, the Poisson assumption gives rise to a characteristic likelihood which is optimised to provide estimates for the parameters in the graduation formula. It has been noted, e.g. page 113 of Gerber (1995), that the same formal expression for the likelihood arises under the different assumption that the central exposures to the risk of death are gamma random variables conditional on the matching numbers of deaths The implications of adopting this dual approach for the parametric graduation process are investigated in this paper. Following Renshaw (1991), both approaches are formulated within the generalised linear modelling (GLM) framework , while the conclusions extend to include non-linear parameterised graduation formulae.

A brief description of the salient features of GLMs is presented in Section 2 for completeness The consequences of switching from the 'conventional' approach to the dual modelling approach when the data are based on head counts, or equivalently, on policy counts in the absence of duplicate police , are discussed in Section 3. The implications for both approaches when duplicate policies are present in the data counts are then discussed in Section 4 and Section 5 respectively Finally an illustration of the implications of the switch from the 'conventional' approach to the dual approach, which reside largely in the reporting of the graduation, is presented in Section 6.

2 GENERALISED LINEAR MODELS

The purpose of this section is to provide a brief introduction to GLMs. A complete treatment of the theory and application can be found in McCullagh & Nelder (1989) and Francis, Green & Payne (1993).

The basis of a GLM is motivated, in the first instance, by the assumption that the data are sampled from a one parameter exponential family of distributions with log-likelihood

$$l = \frac{y\theta - b(\theta)}{\phi} + c(y,\phi)$$

for a single observation y, where θ is the canonical parameter and ϕ is the dispersion parameter, assumed known. It is then straightforward to demonstrate that

$$m = E(Y) = \frac{d}{d\theta}b(\theta) \text{ and } Var(Y) = \phi\frac{d^2}{d\theta^2}b(\theta) = \phi\, b''(\theta).$$

We note that $Var(Y)$ is the product of two quantities The quantity $b''(\theta)$ is called the variance function and depends on the canonical parameter and hence on the mean We can write this as $V(m)$.

The log-likelihoods for some common distributions of interest and which conform to these properties are

$$l = y\log m - m - \log y'$$
$$\theta = \log m, b(\theta) = \exp\theta, V(m) = m, \phi = 1$$

for the Poisson distribution with mean m, and

$$l = \frac{-\dfrac{y}{m} + \log\dfrac{1}{m}}{\dfrac{1}{v}} + v\log y + v\log v - \log\Gamma(v)$$

$$\theta = -\frac{1}{m}, b(\theta) = -\log(-\theta), V(m) = m^2, \phi = v^{-1}$$

for the gamma distribution mean m and variance m^2/v.

More generally a GLM is characterised by independent response variables $\{Y_u\cdot u = 1, 2, \ldots, n\}$ for which

$$E(Y_u) = m_u, Var(Y_u) = \frac{\phi V(m_u)}{\omega_u} \tag{2.1}$$

comprising a variance function V, a scale parameter ($\phi > 0$) and prior weights ω_u.

Covariates enter via a linear predictor

$$\eta_u = \sum_{j=1}^{p} x_{uj}\beta_j$$

with specified structure (x_{uj}) and unknown parameters β_j linked to the mean response through a known differentiable monotonic link function g with

$$g(m_u) = \eta_u.$$

The special link function $g = \theta$, so that $\theta(m) = \eta$, is called the canonical link function. Examples are the log link in the case of the Poisson distribution and the reciprocal link in the case of the gamma distribution

The suffices or units u have structure, either intrinsic or imposed. The data comprise realisations $\{y_u\}$ of the independent response variables, matched to the structure of the units. Generally in any one study, the detail of the distribution and link are fixed, while the predictor structure may be varied

Model fitting is by maximising the quasi log-likelihood

$$q = q(\underline{y}, \underline{m}) = \sum_{u=1}^{n} q_u = \sum_{u=1}^{n} \omega_u \int_{y_u}^{m_u} \frac{y_u - s}{\phi V(s)} ds \qquad (2.2)$$

leading to the system of linear equations

$$\sum_{u=1}^{n} \omega_u \frac{y_u - m_u}{\phi V(m_u)} \frac{\partial m_u}{\partial \beta_j} = 0 \quad \forall j$$

in the unknown $\beta_j s$ These are solved numerically, e.g Francis, Green & Payne (1993), McCullagh & Nelder (1989). Detail of the construction of standard errors for the parameter estimators, based on standard statistical theory, is also to be found in these references Denote the resulting values of the parameter estimators, linear predictor and fitted values, for the current model c, $\hat{\beta}_j, \hat{\eta}_u$ and $\hat{m}_u$ respectively, where

$$\hat{m}_u = g^{-1}(\hat{\eta}_u), \hat{\eta}_u = \sum_{j=1}^{p} x_{uj} \hat{\beta}_j$$

For members of the exponential family of distributions, the quasi log-likelihood is synonymous with log-likelihood The maximal structure possible has the property that the fitted values are equal to the observed responses, that is $\hat{m}_u = y_u$ for all u, and is called the full or saturated model f.

The (unscaled) deviance of the current model c is

$$D(c, f) = d(\underline{y}; \underline{\hat{m}}) = \sum_{u=1}^{n} d_u = \sum_{u=1}^{n} 2\omega_u \int_{\hat{m}_u}^{y_u} \frac{y_u - s}{V(s)} ds = -2 \phi q(\underline{y}; \underline{\hat{m}}),$$

in which the fitted values under the current and saturated models impact on the formula through the lower and upper limits of the integral respectively. The corresponding scaled deviance is

$$S(c, f) = d * (\underline{y}; \underline{\hat{m}}) = \frac{d(\underline{y}, \underline{\hat{m}})}{\phi} = \sum_{u=1}^{n} 2\omega_u \int_{\hat{m}_u}^{y_u} \frac{y_u - s}{\phi V(s)} ds = -2 q(\underline{y}; \underline{\hat{m}}) \qquad (2 3)$$

For fixed distribution, fixed link and hierarchical model structures c_1 and c_2, with c_2 nested in c_1, the difference in scaled deviance

$$S(c_2, f) - S(c_1, f)$$

may be referred, generally as an approximation, to the chi-square distribution with $v_2 - v_1$ degrees-of-freedom, where v_1 and v_2 denote the respective degrees-of-freedom.

Two types of residuals (which are identical only in the case of the Gaussian distribution, for which $V(s) = 1$) are of interest, the Pearson residuals

$$\frac{y_u - \hat{m}_u}{\sqrt{\dfrac{V(\hat{m}_u)}{\omega_u}}} \tag{2.4}$$

or the deviance residuals

$$\text{sign}(y_u - \hat{m}_u)\sqrt{d_u}$$

where d_u is the uth. component of the (unscaled) deviance above

3. HEAD OR POLICY COUNTS WITH NO DUPLICATES

3.1 Distribution Assumptions

In keeping with common practice, let

μ_x = the force of mortality at age x

$_wp_x$ = the probability that a life aged x survives tot age $x + w$

and recall the basic identity

$$_wp_x = \exp - \int_0^w \mu_{x+s}\, ds \tag{3 1}$$

with the implied assumption that μ_x is a function of age alone and is therefore assumed to be constant with respect to variations in calendar time within a fixed observation window.

Focus on a set of individual lives or policyholders. If the latter, and the data are based on policy counts, then it is assumed throughout this Section that all policyholders possess a single policy Individual members of the set are assumed to be observed between ages x and $x + 1$ in the *fixed* calendar period or observation window t to $t + t_0$, with pre-specified policy duration where relevant, and their survival experience is assumed throughout to be independent Typically $t_0 = 4$ years in many United Kingdom (UK) actuarial mortality studies. There is also interest in the case $t_0 = 1$ year when modelling trends in mortality, e g. Renshaw, Haberman & Hatzopoulos (1996). Within such a cell, identified in this instance by the *suffix x*, suppose an individual i enters observation at age v_{xi} and leaves it either by death ($I_{xi} = 1$) or by censorship

$(l_{1j} = 0)$ at age $v_{xi} + w_{xi}$ where $x \leq v_{1i} < v_{1i} + w_{1i} \leq x + 1$. Then it is well known see, e g Section 3.2 of Cox & Oakes (1984), that each such datum contributes an amount

$$L_{1i} = {}_{w_{1i}} p_{v_{1i}} \mu_{v_{1i}+w_{1i}}^{I_{1i}}$$

to the likelihood, or, on resorting to the use of expression (3 1), an amount

$$l_{1i} = \log L_{1i} = - \int_0^{w_{xi}} \mu_{v_{1i}+s} ds + I_{1i} \log \mu_{v_{1i}+w_{xi}}$$

to the log-likelihood. Thus the total contribution to the log-likelihood from such a cell is

$$l_1 = \sum_{i=1}^{n_1} l_{1i} = \sum_{i=1}^{n_x} \left\{ - \int_0^{w_{xi}} \mu_{v_{1i}+s} ds + I_{1i} \log \mu_{v_{1i}+w_{1i}} \right\} \tag{3 2}$$

where the summation extends to all n_1 individuals contributing to the experience in the cell If in addition μ_x is assumed to be piecewise constant with respect to age within each cell and accorded the central value $\mu_{1+1/2}$, expression (3.2) can be written as

$$l_1 = -r_r \mu_{1+1/2} + a_1 \log \mu_{r+1/2}$$

where

$$r_r = \sum_{i=1}^{n_x} w_{1i}, \quad a_x = \sum_{i=1}^{n_x} I_{1i}$$

denote the respective central exposure and actual number of deaths associated with cell x. The expression for the full log-likelihood

$$l = \sum_1 l_1 = \sum_1 \left\{ -r_1 \mu_{1+1/2} + a_x \log \mu_{1+1/2} \right\} \tag{3.3}$$

then follows by summation over all such cells. It is of specific interest to note that this expression may be interpreted in one of two ways

Firstly, and somewhat exclusively in the context of an actuarial graduation, expression (2.3) is identifiable as the kernel of the log-likelihood under the assumption that the actual numbers of deaths, a_1, are modelled as independent realisations of Poisson random variables A_x *conditional* on r_1, such that

$$A_1 \sim \mathrm{Poi}(r_1 \mu_{1+1/2}).$$

For this case, the detail of the distributional requirements to set up the appropriate GLM (equation (2 1) with $u \equiv x$) is either

responses $\{A_x\}$, with $m_1 = r_1 \mu_{1+1/2}, V(m_1) = m_1, \phi = 1, \omega_1 = 1,$ (3 4a)

or equivalently

responses $\{A_1 / r_1\}$, with $m_1 = \mu_{1+1/2}, V(m_1) = m_1, \phi = 1, \omega_1 = r_1$ (3 4b)

Secondly, e.g. Section 11 5 of Gerber (1995), expression (3 3) is also identifiable as the kernel of the log-likelihood under the assumption that the exposures to risk, r_x, are modelled as independent realisations of gamma random variables R_x *conditional* on a_x, such that

$$R_x \sim \text{gam}(a_x, \mu_{x+1/2})$$

Superficially this result is perhaps a little unusual in-so-far as the gamma distribution is generally associated with two unknown parameters, whereas here, as with the Poisson distribution above, there is only a single parameter to estimate For this case, the detail of the distributional requirements to set up the appropriate GLM (equation (2 1) with $u \equiv x$) is either

$$\text{responses } \{R_x\}, \text{ with } m_x = a_x \frac{1}{\mu_{x+1/2}}, V(m_x) = m_x^2, \phi = 1, \omega_x = a_x, \qquad (3\ 5a)$$

or equivalently

$$\text{responses } \{R_x / a_x\}, \text{ with } m_x = \frac{1}{\mu_{x+1/2}}, V(m_x) = m_x^2, \phi = 1, \omega_x = a_x \qquad (3.5b)$$

The data comprise the ordered pairs of numbers of deaths and central exposures (a_x, r_x) over a range of ages x All of the r_xs are non-zero by implication, but it is conceivable that certain of the $a_x s$ are zero. This is most likely to occur at the extremities of the age range were the data are sometimes sparse Note that while such data cells are retained in any analysis of the data based on distributional assumptions (3.4a & b), they are weighted out of any analysis based on distributional assumptions (3.5a & b)

3.2 Discussion

The optimisation of expression (3 3) under the former interpretation (based on the Poisson distribution) is central to the current graduation practice of the Continuous Mortality Investigation (CMI) Bureau in the UK, e g. Forfar *et al* (1988), while the optimisation of expression (3.3) under the alternative interpretation (based on the gamma distribution) would appear not to have been investigated previously in an actuarial graduation setting.

It is possible to derive the first set of assumptions, in which the number of actual deaths A_x form the response variables, by taking expectations and variances under the identity

$$A_x = \sum_{i=1}^{n_x} I_{xi}$$

where I_{xi} is the zero-one indicator random variable, introduced previously, in Section 3.1. It has the property

$$E(I_{xi}) = E(I_{xi}^2) = P(I_{xi} = 1) = 1 - \exp - \int_0^{n_{xi}} \mu_{x+s} ds$$

and is assumed to be independent for all individuals i. The results then follow under the assumption that μ_x is piecewise constant within cells, so that

$$E(I_{xi}) = (E(I_{xi}^2) = 1 - \exp(-\mu_{x+1/2}w_{xi}),$$ (3 6)

and on neglecting second and higher order therms in the power series expansion of $\exp(-\mu_{x+1/2}w_{xi})$, so that

$$\text{Var}(I_{xi}) = E(I_{xi}) \approx \mu_{x+1/2}w_{xi}.$$

Under the second set of assumptions, for which the responses satisfy

$$R_x = \sum_{i=1}^{n_x} W_{xi},$$

the individual exposures W_{xi} are modelled as random variables. Under the additional assumption that the individual exposures are independent and identically distributed, it follows trivially from the reproductive property of the gamma distribution that they have the gamma distribution

$$W_{xi} \sim \text{gam}(\frac{a_x}{n_x}, \mu_{x+1/2})$$

Again based on the reproductive property of the gamma distribution, note that it is also possible to construct the identical GLM by defining

$$R_x = \sum_{i=1}^{n_x} W_{xi} = \sum_{j=1}^{a_x} T_{xj}$$

in which the T_{xj}s are assumed to be independent and identically distributed gamma random variables, such that

$$T_{xj} \sim \text{gam}(1, \mu_{x+1/2}),$$

and where at least one death is recorded in every cell Here it is possible to interpret T_{xj} as the sum of randomly selected censored exposures W_{xi} the last of which is associated with a death

The target of the graduation process is the *force of mortality* μ_x under distribution assumptions (3.4a & b) and the *force of vitality* $1/\mu_x$ under distribution assumptions (3.4a & b). In using the latter description, we follow the terminology of Lambert (1772) see, e g Daw (1980)

The value of the scaled deviance, (expression 2.3, with $u \equiv x$) is identical under both sets of modelling assumptions (3 4a & b) and (3.5a & b) and is equal to

$$S(c, f) = \sum_x 2\left\{ a_x \log\frac{a_x}{r_x\hat{\mu}_{x+1/2}} - (a_x - r_x\hat{\mu}_{x+1/2}) \right\}$$ (3 7)

where $\hat{\mu}_x$ denotes the graduated values of μ_x provided deaths are recorded for all ages (i e. $a_x > 0 \; \forall \; x$) so that none of the terms are weighted out of the expression on the right hand side (RHS) of equation (3 7) under the dual modelling assumptions (3.5a &

b) This is perhaps a surprising result on the surface It reflects the fact that the same objective function, expression (3 3), which is embedded in the construction of the scaled deviance as the quasi log-likelihood function, (expression 2 2, with $u \equiv x$) is optimised when fitting the model structure (or graduation formula).

Subject to the weighting out of any data cells containing zero $a_x s$ in the one case, the two sets of distribution assumptions lead to identical graduations for μ_x Thus, assumption (3 4a) with responses $\{a_x\}$ in combination with log-link based graduation formulae of the type

$$\log \mu_{x+1/2} = \sum_{j=0}^{p} h_{xj} \beta_j \qquad (3 8)$$

so that

$$\log m_x = \eta_x = \log r_x + \log \mu_{x+1/2} = \log r_x + \sum_{j=0}^{p} h_{xj} \beta_j ,$$

gives identical graduations to those obtained under assumption (3 5b) with responses $\{r_x\}$ so that

$$\log m_x = \eta_x = \log a_x - \log \mu_{x+1/2} = \log a_x + \sum_{j=0}^{p} h_{xj} \beta_j .$$

Typically the parameterised structure of the RHS of the graduation equation (3.8) is a polynomial in x with either the log r_x or log a_x terms declared as offsets, as the case may be The estimated values of the parameters β_j are identical in magnitude but opposite in sign in the two cases Similarly assumption (3.4b) with responses $\{a_x/r_x\}$ in combination with the power link graduation formulae of the type

$$\mu_{x+1/2}^{\gamma} = \sum_{j=0}^{p} h_{xj} \beta_j$$

gives identical graduations to those obtained under assumption (3 5b) with responses $\{r_x/a_x\}$ so that

$$\mu_{x+1/2}^{-\gamma} = \sum_{j=0}^{p} h_{xj} \beta_j$$

This time the estimated values of the parameters β_j are identical in both magnitude and sign in the two cases. Thus the general conclusions of this paper extend to non-linear parameterised graduation formulae via the identity link under the 'conventional' approach and the reciprocal link under the dual approach.

Let $e_x = r_x \mu_{x+1/2}$ denote the expected number of deaths predicted at age x, under the conventional graduation methodology encapsulated by equations (3.4a &b), and define the statistics

$$dev_x = a_x - e_x, \sqrt{V_x} = \sqrt{e_x}, zx = \frac{dev_x}{\sqrt{V_x}}, \ 100\frac{a_x}{e_x}. \tag{3.9}$$

It is common practice for these to be tabulated (subject to possible cell grouping in the tails of the age range) as part of the diagnostic checking procedure of a graduation. Note in particular that the statistic z_x is the Pearson residual of the corresponding GLM, (expression 2.3, with $u \equiv x$). Thus typically the value of the approximate chi-square statistic $\sum_x z_x^2$ is quoted as one of the many test statistics of a graduation. The equivalent statistics under the dual graduation methodology encapsulated by equations (3.5a or b) involving definition $\tilde{e}_x = a_x / \hat{\mu}_{x+1/2}$ or expected exposure predicted at age x, are

$$\tilde{dev}_x = r_x - \tilde{e}_x, \sqrt{\tilde{V}_x} = \sqrt{\frac{\tilde{e}_x^2}{a_x}}, \tilde{z}_x = \frac{\tilde{dev}_x}{\sqrt{\tilde{V}_x}}, \ 100\frac{r_x}{\tilde{e}_x}. \tag{3 10}$$

Again note that these statistics are defined in such a way that $\tilde{z}_x$ denotes the Pearson residual of the associated GLM (3.5a or b). The relationship between the values of the deviation under the dual and 'conventional' graduation methodologies, namely

$$\tilde{dev}_x = \frac{-\tilde{dev}_x}{\hat{\mu}_{x+1/2}}$$

implies that the residuals under the two methodologies have opposite signs. Although only strictly exact provided all the a_xs are positive, this relationship provides a very close approximation when the a_xs take zero values at the extremities of the age range concerned. Detailed examination of the respective formulae defining the Pearson residuals z_x and $\tilde{z}_x$ reveals that they differ in magnitude (and have opposite signs) On the other hand, because of the equality of the deviance components under the two methodologies established above, the deviance residuals defined by either

$$\text{sign}(dev_x)\sqrt{d_x} \text{ or sign}(\tilde{dev}_x)\sqrt{d_x}$$

as the case may be, where d_x is the general term in the summation on the RHS of expression (3 7), are identical in magnitude (and opposite in sign) under the dual methodologies. It is also of interest to note that the final statistics quoted in expressions (3 9) and (3 10), corresponding to the respective dual modelling scenarios, are the reciprocals of one another prior to scaling by 100 Again both of these features are exact when all the a_xs are positive and represent a very close approximation when any of the a_xs are zero at the extremities of the age range

4 POLICY COUNTS WITH DUPLICATES: CLAIM NUMBER RESPONSE MODELS

4.1 Preliminaries

The data used in the construction of actuarial life tables are generally based on policy rather than head counts Consequently, the death of a policyholder with more than one policy will appear as more than one death in the raw data The resulting graduation needs to account for this overdispersion. for a review of the issues involved, readers should consult Forfar *et al.* (1988) and Renshaw (1992).

Let

D_{xi} = the number of policies held by policyholder i, age x

C_{xi} = the number of policies held by policyholder i, age x, resulting in a claim.

Assume that the random variables D_{xi} are i.i.d $\forall\, i$ and let D_x denote the generic type. For each i, the events $(C_{xi} = k \mid I_{xi} = 1)$ and $(D_{xi} = k)$ are such that

$$(C_{xi} = k \mid I_{xi} = 1) \Leftrightarrow (D_{xi} = k), k = 1, 2, 3, ..$$

and thus have identical probabilities. Define

$$P(D_x = k) = P(C_{xi} = k \mid I_{xi} = 1) = \begin{cases} \pi_x^{(k)} & k = 1, 2, 3, ... \\ 0 & \text{otherwise} \end{cases}$$

where

$$\pi_x^{(k)} \geq 0, \sum_{k=1}^{\infty} \pi_x^{(k)} = 1$$

Denote

$$E(D_x) = E(C_{xi} \mid I_{xi} = 1) = \sum_{k=1}^{\infty} k\pi_x^{(k)} =_1 \pi_x$$

and

$$E(D_x^2) = E(C_{xi}^2 \mid I_{xi} = 1) = \sum_{k=1}^{\infty} k^2 \pi_x^{(k)} =_2 \pi_x.$$

It also follows by definition that

$$P(C_{xi} = 0 \mid I_{xi} = 0) = 1$$

so that

$$E(C_{xi} \mid I_{xi} = 0) = E(C_{xi}^2 \mid I_{xi} = 0) = 0$$

Hence the *unconditional* distribution of C_{xi} is given by

$$P(C_{xi} = k) = \begin{cases} 1 - E(I_{xi}), & k = 0 \\ E(I_{xi})\pi_x^{(k)}, & k = 1, 2, 3, \end{cases}$$

for which

$$E(C_{\,\iota\iota})={}_1\pi_{\,\chi}E(I_{\,\iota\iota}), \; E(C_{\iota\iota}^2)={}_2\pi_{\,\lambda}E(I_{\,\iota\iota})$$

These equations, in combination with expression (3 6) for $E(I_{\,\iota\iota})$, on neglecting second and higher order terms in the power series expansion of $\exp(-\mu_{x+1/2}\,w_i)$, imply that

$$E(C_{\,\lambda\iota})\approx{}_1\pi_{\,\iota}\mu_{\,\lambda+1/2}w_{\,\iota\iota} \; \text{and} \; \text{Var}(C_{\,\iota\iota})\approx{}_2\pi_{\,\iota}\mu_{\,\iota+1/2}w_{\,\iota\iota}. \qquad (4.1)$$

We also have an interest in the first two moments of the product random variable $D_{\iota\iota} I_{\iota\iota}$. Under the mild assumption that the number of policies, $D_{\iota\iota}$, held by policyholder ι, aged x, is statistically independent of the mode of censorship, $I_{\iota\iota}$, it follows that

$$E(D_{\iota\iota}I_{\iota\iota}) = E(D_{\iota\iota})E(I_{\iota\iota}), \; \text{Var}(D_{\iota\iota}I_{\iota\iota}) = E(D_{\iota\iota}^2)E(I_{\lambda\iota}^2) - \{E(D_{\iota\iota})E(I_{\iota\iota})\}^2$$

These equations in combination with expressions (3 6), on neglecting second and higher order terms in the power series expansion of $\exp(-\mu_{x+1/2}\,w_i)$, then imply that

$$E(D_{\iota\iota}I_{\iota\iota})\approx{}_1\pi_{\,\iota}\mu_{\,\iota+1/2}w_{\,\iota\iota} \; \text{and} \; \text{Var}(D_{\iota\iota}I_{\iota\iota})\approx{}_2\pi_{\,\iota}\mu_{\,\iota+1/2}w_{\,\iota\iota} \qquad (4.2)$$

4.2   Distribution Assumptions

Let

$A'_{\,\iota}$ = the number of policies giving rise to a claim through deaths

$r'_{\,\iota}$ = the cental exposure to the risk of death based on policies.

Note that
$$r'_{\,\iota} = \sum_{\iota=1}^{n_{\,\iota}} d_{\,\iota\iota} w_{\,\lambda\iota}$$

where $d_{\,\iota\iota}$ (≥ 1) denotes the number of policies held by policyholder ι, reducing to $r_{\,\lambda}$ if and only if $d_{\,\iota\iota} = 1 \; \forall \; i$. Throughout this Section the A'_xs are modelled as random variables *conditional* on $r'_{\,\iota}$. It follows on taking expectations and variances under any one of the following identities

$$A'_{\,\iota} = \sum_{\iota=1}^{A_{\,\iota}} D_{\,\iota\iota} \; \text{(with } A_{\,\iota} > 0\text{)}, \; A'_{\,\iota} = \sum_{\iota=1}^{n_{\,\iota}} C_{\,\lambda\iota}, \; A'_{\,\iota} = \sum_{\iota=1}^{n_{\,\iota}} D_{\,\iota\iota} \, I_{\,\iota\iota} \qquad (4\,3)$$

that the detail of the distributional requirements to set up the appropriate GLM (equation (2.1), with $u \equiv x$) is either

$$\text{responses } \{A'_{\,\iota}\}, \text{ with } m_{\,\iota} = r'_{\,\iota}\mu_{\,\iota+1/2}, \, V(m_{\,\iota}) = m_{\,x}, \, \phi = 1, \, \omega_x = \phi_{\,\iota}^{-1}, \qquad (4\,4a)$$

or equivalently

$$\text{responses } \{A'_{\,\iota} \, / \, r'_{\,\iota}\}, \text{ with } m_{\,\lambda} = \mu_{\,\lambda+1/2}, \, V(m_{\,\iota}) = m_{\,\iota}, \, \phi = 1, \, \omega_{\,\iota} = r'_{\,\iota}\phi_{\,\iota}^{-1}, \qquad (4\,4b)$$

where
$$\phi_{\,\lambda} = \frac{{}_2\pi_{\,\iota}}{{}_1\pi_{\,\lambda}}$$

4.3 Discussion

The result (4.4a) follows from the first of the identities (4.3) which, under the assumption that A_x is independent of the $\{D_u\}$ implies, in combination with equations (3.4a)

$$E(A'_x) = E(D_x)E(A_x) = {}_1\pi_x r_x \mu_{x+1/2}$$

and

$$\text{Var}(A'_x) = \text{Var}(D_x)E(A_x) + \{E(D_x)\}^2 \text{Var}(A_x) = \frac{E(D_x^2)}{E(D_x)} E(A'_x) = \frac{2\pi_x}{{}_1\pi_x} E(A'_x)$$

Under the independence of the terms in the respective summations, the same result follows trivially from either the second of the identities (4.3) in combination with equations (4 1), or the third of the identities (4.3) in combination with equations (4.2). In all three cases, the product term ${}_1\pi_x r_x$ in the expression for $E(A'_x)$ involving the unobserved central exposure based on lives has been replaced by r'_x, the observed central exposure based on policies The result (4.4b) follows trivially from result (4 4a)

The justification for (4.4a) based on the second of the identities (4.3) and equations (4.1) is a generalisation of the method described in Renshaw (1992) for initial exposures and the binomial response model. This work establishes a link with much earlier work on the modelling of duplicate policies using an empirical approach, e.g. Beard & Perks (1949).

A knowledge of the reciprocals of the overdispersion parameters ϕ_x is needed to form the weights, if the distributional assumptions (4 4) are to be fully implemented Insight into the potential variation of ϕ_x with x is provided by studies of the properties of so-called variance ratios, the empirical equivalent of ϕ_x, e.g Forfar *et al.* (1988). These are defined as

$$vr_x = \frac{\sum_i i^2 f_x^{(i)}}{\sum_i if_x^{(i)}}$$

where $f_x^{(i)}$ denotes the proportion, at age x, of policyholders who have i policies and where

$$f_x^{(i)} \geq 0 \ \forall \ i = 1, 2, 3, \ldots , \quad \sum_i if_x^{(i)} = 1 \Rightarrow vr_x \geq 1$$

There are a number of alternative practical possibilities When available, variance ratios can be used as estimates for the dispersion parameters ϕ_x and graduation can proceed in accordance with assumptions (4 4) On the other hand, Forfar *et al.* (1988) acting for the CMI Bureau in the UK, elect to transform the data by dividing both the policy counts a'_x and exposures r'_x by the matching variance ratios prior to graduation with assumptions (3 4) displacing assumptions (4 4) When a detailed knowledge of the relevant variance ratios is not available for analysis a possible method of generating estimates for the dispersion parameters is described in Renshaw (1992). Alterna-

tively, under the assumption that the underlying modelling distribution of the number of duplicate policies is identical across all ages x in the absence of any further detailed knowledge about this distribution, the dispersion parameters ϕ_x may be replaced by a constant scale (or dispersion) parameter ϕ in assumptions (4 4), e.g. Renshaw (1992) It is estimated as

$$\hat{\phi} = \frac{\text{unscaled deviance}}{\text{degrees - of - freedom}}$$

and is root $\sqrt{\hat{\phi}}$ used to scale the Pearson residuals z_x of expressions (3 9) or $\tilde{z}_x$ of expressions (3 10), by multiplying either V_x or $\tilde{V}_x$ by $\hat{\phi}$, as the case may be. Here the unscaled deviance is calculated using the expression on the RHS of equation (3.7). (Recall that ϕ was set to one when deriving this expression, so that the scaled deviance $S(c, f)$ is also the unscaled deviance in this instance.) This latter approach is closest in spirit to that adopted by Forfar *et al.* (1988) involving the transformation of the data prior to graduation in-so-far as it produces identical graduations, while allowing the presence of duplicate policies to impact solely on the second moment properties of the graduation process

5. POLICY COUNTS WITH DUPLICATES: EXPOSURE RESPONSE MODELS

5.1 Preliminaries

As before, let

D_{xi} = the number of policies held by policyholder i, age x
W_{xi} = the contribution to the exposure by policyholder i, age x

Recall that D_x, D_{xi} are assumed to be i.i d. $\forall$ i with

$$E(D_x) = {}_1\pi_x, \; E(D_x^2) = {}_2\pi_x.$$

Recall also the duality property of Section 3 2, namely that the central exposure to risk of death based on head counts, at age x

$$R_x = \sum_{i=1}^{n_x} W_{xi} \sim \text{gam}(a_x, \mu_{x+1/2}),$$

so that

$$E(R_x) = \frac{a_x}{\mu_{x+1/2}}, \; E(R_x^2) = \frac{a_x(1 + a_x)}{\mu_{x+1/2}^2}.$$

Consider the identity

$$R'_{\chi} = \sum_{i=1}^{n_{\chi}} D_{\chi i} W_{\chi i} \tag{5.1}$$

which defines the central exposure to risk of death based on policy counts, at age x Assuming that the number of policies held by an individual policyholder is independent of the corresponding contribution to the exposure to risk from that individual and that the individual exposures are independent, it follows from the identity (5 1) that

$$E(R'_{\chi}) = E(D_{\chi}) \sum_{i=1}^{n_{\chi}} E(W_{\chi i}) = E(D_{\chi})E(R_{\chi}) = \frac{{}_{1}\pi_{\chi}a_{\chi}}{\mu_{\chi+1/2}} \tag{5.2}$$

and

$$E(R_{\chi}'^{2}) = E(D_{\chi}^{2})E(\sum_{i=1}^{n_{\chi}} W_{\chi i})^{2} = E(D_{\chi}^{2})E(R_{\chi}^{2}) = \frac{2\,\pi_{\chi}a_{\chi}(1+a_{\chi})}{\mu_{\chi+1/2}^{2}} \tag{5 3}$$

after simplification.

5.2 Distribution Assumptions

Let

R'_{χ} = the central exposure to the risk of death based on policies
a'_{χ} = the number of policies giving rise to a claim through deaths

Throughout this section the R'_{χ} s are modelled as random variables *conditional* on a'_{χ}. It follows from equations (5.1), (5.2) and (5.3) that the detail of the distributional requirements to set up the appropriate GLM (equation (2 1), with $u \equiv x$) is either

$$\text{responses } \{R'_{\chi}\}, \text{ with } m_{\chi} = a'_{\chi} \frac{1}{\mu_{\chi+1/2}}, \; V(m_{\chi}) = m_{\chi}^{2}, \phi = 1, \omega_{\chi} = \psi_{\chi}^{-1}, \tag{5 4a}$$

or equivalently

$$\text{responses } \{R'_{\chi} / a'_{\chi}\}, \text{ with } m_{\chi} = \frac{1}{\mu_{\chi+1/2}}, \; V(m_{\chi}) = m_{\chi}^{2}, \phi = 1, \omega_{\chi} = \psi_{\chi}^{-1}, \tag{5.4b}$$

where this time

$$\psi_{\chi} = \left(\frac{2\pi_{\chi}}{{}_{1}\pi_{\chi}^{2}} - 1 \right) + \frac{2\pi_{\chi}}{{}_{1}\pi_{\chi}} \frac{1}{a'_{\chi}}. \tag{5 5}$$

5.3 Discussion

In parallel with the previous case, this time the product term $_1\pi_x a_x$ in the expression for $E(R'_x)$ involving the unobserved number of deaths a_x based on head counts has been replaced by a'_x, the observed number of deaths base on policy counts. Again result (5 4b) follows trivially from result (5 4a)

A knowledge of the reciprocals of the dispersion parameters ψ_x is required to form the weights if the distribution assumptions (5.4a or b) are to be fully implemented In the event that the results of a study into the variance ratios for the policies in question are available, this will furnish estimates for the first two moments $_1\pi_x$ and $_2\pi_x$ of the number of duplicate policies so that modelling can proceed. Alternatively if it is assumed that the square of the coefficient of variation of the number of duplicate policies held by an individual is sufficiently small so as to make the first term on the RHS of expression (5.5) for ψ_x is negligible in comparison with the second term,

$$\psi_\lambda = \phi_x \frac{1}{a'_\lambda}$$

and the situation is analogous to that discussed in Section 4.3.

6. ILLUSTRATION

The dual methodologies are illustrated using the Pensioners' widows 1979-1982 experience reported in Table 15.5 of Forfar *et al* (1988). The data (a_x, r_x), comprising the numbers of deaths a_x and matching central exposures r_x, are reported in the age range 17 to 108 years inclusive. There are $2 + 5 = 7$ completely empty cells in the extremities of the age range and $28 + 12 = 40$ cells contain no reported deaths The detail of the graduation contained in the above Table is based on Gompertz's formula fitted by the 'conventional' approach, in which the numbers of deaths are modelled as Poisson random variables The data have been regraduated using both the 'conventional' approach based on assumptions (3.4a) with predictor-link formulation

$$\log m_x = \log r_x + \log \mu_{x+1/2} = \log r_x + \beta_0 + \beta_1 \left(\frac{x - 70}{50} \right),$$

and the dual approach based on assumptions (3 5a) with equivalent predictor-link formulation

$$\log m_x = \log a_x - \log \mu_{x+1/2} = \log a_x + \beta_0 + \beta_1 \left(\frac{x - 70}{50} \right),$$

where m_x denotes the respective mean responses The associated graduation formula, implied by these formulae, is taken from Forfar *et al* (1988). Some details of the respective fits including the parameter estimates are recorded in Table 6.1 The corresponding parameter estimates have opposite signs as expected, but differ slightly in absolute value because the data entries involving zero deaths feature only in the 'conventional' analysis Similarly the corresponding values of both the deviances and

the degrees-of-freedom differ for the same reason. These differences are found to disappear when the 'conventional' analysis is applied to the reduced data set and identical graduations result as a consequence (subject to very minor differences induced by the numerical fitting algorithm operating under the two different approaches) An extract of both graduations based on the detail of Table 6 1 is reproduced in Table 6 2(a&b), along with detail of the associated statistics of expressions (3 9) and (3.10), as the case may be The detail of Table 6.2a is in complete agreement with that to be found in Table 15.5 of Forfar *et al* (1988), while the relatively minor effects of the excluded data under the dual modelling approach are demonstrated. The basic differences in the accompanying statistics used to monitor the effectiveness of a graduation under the two different approaches, as described in Section 3.2, can be verified

7 CONCLUSIONS

The 'conventional' actuarial approach to the construction of μ_x-graduations based on the fitting of a wide class of parameterised mathematical formulae by optimising the likelihood, in which the death counts are modelled as Poisson random variables *conditional* on the central exposures, is effectively equivalent to a dual approach in which the central exposures are modelled as gamma random variables *conditional* on the death counts The dual approaches lead to identical graduations provided deaths are recorded in all data cells, otherwise small differences occur in practice as a consequence of the loss of information from any data cells in which no deaths are recorded under the one approach Key differences occur in the diagnostic statistics of a graduation, with residuals being accorded opposite signs under the two different approaches In practice, a detailed knowledge of the specific nature of the empirical distributions on duplicate policies has only a minimal effect on the first moment of a graduation under the two formulations described here In the absence of this knowledge, these first moment properties may be neglected and a free standing constant scale (or dispersion) parameter introduced, under either formulation, to represent the second moment properties of a graduation in the presence of duplicate policies.

   The dual approach to μ_x-graduation would appear to have distinct advantages over the 'conventional' approach to graduation, when it is adapted and applied to the construction of select mortality tables. This is discussed further in Renshaw & Haberman (1996), who successfully use the dual approach to model the log crude mortality ratios for individual select durations relative to the ultimate experience

REFERENCES

BFARD, R E & PERKS, W (1949) The Relation between the Distribution of Sickness and the Effect of Duplicates on the Distribution of Deaths *J I A* , 75, 75
COX, D R & OAKES, D (1984) Analysis of Survival Data *Chapman and Hall*
DAW, R H (1980) Johann Heinrich Lambert (1728-1777) *J I A* , 107, 345-363
FRANCIS, B , GREEN, M & PAYNE, C (1993) Eds The GLIM System Release 4 Manual *Clarendon Press Oxford*

FORFAR, D O , MCCUTCHEON, J J & WILKIE, A D (1988) On Graduation by Mathematical Formula *J I A* , 115 , 1-135 and *T F A* , 41, 97

GERBER, H U (1995) Life Insurance Mathematics (2nd edition) *Springer*

LAMBERT, J H (1772) Anmerkungen uber die Sterbluhkeit, Todtenlisten, Gebusthen un Ehen Vol III, 475-599

MCCULLAGH, P & NELDER, J A (1989) Generalized Linear Models (2nd edition) *Chapman & Hall*

RENSHAW, A E (1991) Actuarial Graduation Practice and Generalized Linear & Non-Linear Models *J I A* , 118, 295-312

RENSHAW, A E (1992) Joint Modelling for Actuarial Graduation and Duplicate Policies *J I A* , 119, 69-85

RENSHAW, A E & HABERMAN, S (1996) Dual Modelling and Select Mortality *I M E To appear*

RENSHAW, A E , HABERMAN, S & HATZOPOULOS, P (1996) Recent Mortality Trends in U K Male Assured Lives *B A J* , 2, 449-477

TABLE 6 1

PARAMETERS ESTIMATES WITH (STANDARD ERRORS)

'conventional' approach	dual approach
deviance is 60 98 with 83 d f	deviance is 45 99 with 50 d f
scale parameter $\phi=1$	scale parameter $\phi=1$
$\hat{\beta}_0 = -3\ 553\ (0\ 03923)$	$\hat{\beta}_0 = 3\ 543\ (0\ 03925)$
$\hat{\beta}_0 = 4\ 317\ (0\ 1966)$	$\hat{\beta}_0 = -4\ 332\ (0\ 1979)$

TABLE 6 2(a)

GRADUATION EXTRACT, 'CONVENTIONAL' METHOD

x	r_x	$\mu_{x+1/2}$	a_x	e_x	dev_x	$\sqrt{V_x}$	z_x	$100a_x/e_x$
17	0 5	0 00029	0	0 00	0 00	-	-	-
30	36 0	0 00091	0	0 03	-0 03	-	-	-
40	115 5	0 00215	0	0 25	-0 25	-	-	-
50	378 5	0 00509	3	1 93	1 07	-	-	-
60	1029 0	0 01208	14	12 43	1 57	3 53	0 45	112 6
65	1029 0	0 01860	21	19 14	1 86	4 37	0 43	109 7
70	941 0	0 02864	21	26 95	-5 95	5 19	-1 14	77 9
75	607 0	0 04410	33	26 77	6 23	5 17	1 20	123 3
80	323 5	0 06790	25	21 97	3 03	4 69	0 65	113 8
85	132 5	0 10455	11	13 85	11 60	3 72	-0 77	79 4
95	4 0	0 24790	2	0 99	1 01	-	-	-
108	2 0	0 76154	0	1 52	-1 52	-	-	-

TABLE 6.2(b)

GRADUATION EXTRACT, DUAL METHOD

x	a_x	$\mu_{x+1/2}$	r_x	$\tilde{e}_x$	$d\tilde{e}v_x$	$\sqrt{\tilde{V}_x}$	$\tilde{z}_x$	$100r_x/\tilde{e}_x$
17	0	0 00029	0 5	:				
30	0	0 00090	36 0					
40	0	0 00215	115 0					
50	3	0 00511	378 5	586 7	-208 2	338 7	-0 61	64 5
60	14	0 01216	1029 0	1151 1	-122 1	307 6	-0 40	89 4
65	21	0 01876	1029 0	1119 6	-90 6	244 3	-0 37	91 9
70	21	0 02893	941 0	725 9	215 1	158 4	1 36	129 6
75	33	0 04461	607 0	739 7	-132 7	128 8	-1 03	82 1
80	25	0 06880	323 5	363 4	-39 9	72 7	-0 55	89 0
85	11	0 10611	132 5	103 7	28 8	31 3	0 92	127 8
95	2	0 25237	4 0	7 9	-3 9	5 6	-0 70	50 5
108	0	0 77841	2 0					