JIA 118 (1991) 295-312

J.I.A. **118**, II, 295–312

ACTUARIAL GRADUATION PRACTICE AND GENERALISED LINEAR AND NON-LINEAR MODELS

By A. E. Renshaw, B.Sc., Ph.D.
(*of The City University, London*)

ABSTRACT

It is demonstrated how existing actuarial graduation practice, used in the construction of life tables, can be extended to considerable effect by formulating the techniques within the generalised linear and non-linear modelling framework.

KEYWORDS

Graduation; Models; Life Tables

1. INTRODUCTION

Many of the models used by actuaries for graduation purposes, both historical and contemporary, are specific instances of the rich class of generalised linear models (GLMs); a potentially unifying feature which would appear to have gone largely unobserved in actuarial literature. The aims of this paper, therefore, are to highlight these connections where they exist, and to exploit the benefits which accrue from adopting this different perspective on largely existing actuarial practice. One immediate practical benefit derives from the exploitation of the attendant computer software package (GLIM) as a means of implementing graduations. Scrutiny of residual plots as an additional diagnostic check on any adopted graduations, which would not appear to be current practice, is also advocated to supplement the more formal statistical tests in common use.

Much of what follows is motivated by the recently published comprehensive paper on graduation by Forfar, McCutcheon & Wilkie (1988) (together with the attendant *Continuous Mortality Investigation* (*CMI*) *Report* No. 9), and, to a lesser extent, by relevant sections of the standard textbook by Benjamin & Pollard (1980) (in particular Chapter 14). A concerted effort has been made to adopt the notation of the former reference where appropriate, since this contribution is essentially supplementary to that paper.

Sections 2 and 3 contain a brief account of the salient features of a GLM together with an indication of the more obvious features likely to be of immediate interest in graduation. Section 4 indicates how existing actuarial graduation practice fits within the GLM context, providing both a more unified modelling framework within which to work, as well as leading to alternative models for possible consideration. The reanalysis of various published graduations using GLM techniques and graphical diagnostics is discussed in Section 5. Section 6 is devoted to generalised non-linear models (GNLMs), and to how these might be implemented. There is a brief discussion of a reanalysis of some published graduations in Section 7.

2. GENERALISED LINEAR MODELS (GLMS)

Focus attention on a (column) vector of responses $\boldsymbol{y}=(y_i)$ $i=1, 2, \ldots, n$ treated as a realisation of a vector of independent random variable $\boldsymbol{Y}=(Y_i)$ with systematic deterministic structure defined through the vector of means $\boldsymbol{m}=E(\boldsymbol{Y})$. (Notationally m rather than μ is used here to denote the mean, so that the latter may be reserved to denote the force of mortality in the sequel.) Within this context a GLM is characterised by three basic ingredients comprising:

(1) a modelling distribution imparted to the independent random variable, Y_i,
(2) the inclusion of model covariates $\boldsymbol{x}_j$ through a linear predictor:

$$\boldsymbol{\eta}=\sum_{j=1}^{p}\boldsymbol{x}_j\beta_j \qquad \text{which is}$$

(3) linked into the model through the means $\boldsymbol{m}$ by the so-called link function g where:

$$\boldsymbol{\eta}=g(\boldsymbol{m}).$$

As a general rule the modelling distributions available for use are restricted to the exponential family of distributions, which includes the normal, binomial and Poisson distributions amongst its members. The linear predictor:

$$\eta_i=\sum_{j=1}^{p}x_{ij}\beta_j \qquad i=1, 2, \ldots, n,$$

comprises a known covariate structure $\boldsymbol{X}=(x_{ij})=(\boldsymbol{x}_j)$, a $n\times p$ matrix with columns $\boldsymbol{x}_j$. The parameters β_j are usually, but not always all, unknown. Precisely what constitutes the linearity of such a predictor will be discussed in context later. The link function, g, is monotonic—so that its inverse g^{-1} exists—and is defined to be differentiable over its domain.

Model fitting is done using the GLIM computer package, in which the βs are estimated by maximum likelihood using an iterative weighted least squares algorithm. The reader is referred to the text book by McCullagh & Nelder (1989) for further detail.

3. GRADUATION AND GLMS

3.1 *Modelling Distributions*

The crude data to be smoothed by graduation comprise the number of recorded deaths, A_x, accruing from exposures, R_x, over a range of ages x. The target is either the probability of death, q_x, based on initial exposures or the force of mortality, μ_x, based on central exposures. It is not intended to target the central rate of mortality, as this involves no new detail. Define, therefore, the responses of the GLMs to be the numbers of deaths, A_x, at age x. Here $i=x$ and $Y_i=A_x$. Further, in common with Forfar *et al.* (1988), when targetting q_x use the binomial modelling distribution:

$$A_x \sim B(R_x, q_x)$$

independently for all x, with mean:

$$m_x = R_x q_x;$$

and when targetting μ_x use the Poisson modelling distribution:

$$A_x \sim P(R_x \mu_{x+\frac{1}{2}})$$

independently for all x, with mean:

$$m_x = R_x \mu_{x+\frac{1}{2}}.$$

3.2 *Linear and Non-Linear Predictors*

In common with the familiar graduation scenario, concentrate on homogeneous data sets of the type (R_x, A_x) for which age x is deemed to be the only covariate affecting death. Then any polynomial in x defines an obvious linear predictor:

$$\eta_x = \beta_0 + \beta_1 x + \ \ldots\ \beta_{s-1} x^{s-1}. \tag{3.1}$$

Perhaps a written word about the linearity of the predictor is warranted at this stage. The ages x are known and regression coefficients β_j unknown. Thus, for a polynomial predictor of fixed degree $s-1$, the predictor changes according to the choice of βs. Let η_x' and η_x'' denote predictors based on two such choices, in which the βs of equation 3.1 are replaced respectively by β's and β''s. Then the predictor structure defined by equation 3.1 is deemed to be linear, by virtue of the fact that any linear combination:

$$\eta_x^* = a'\eta_x' + a''\eta_x'' \tag{3.2}$$

for any scalars a' and a'' can also be written in the same form as equation 3.1. In fact, in this instance, the βs of equation 3.1 are replaced by:

$$\beta_j^* = a'\beta_j' + a''\beta_j''.$$

By the same token, predictors containing one or more exponential terms, such as the familiar Makeham form:

$$\eta_x = A + Bc^x, \tag{3.3}$$

are non-linear, by virtue of the fact that η_x^*, constructed in a similar way but using equation 3.3 in conjunction with equation 3.2, cannot be expressed in the same form as equation 3.3.

The Forfar *et al.* (1988) paper makes extensive use of predictors based on the so-called Gompertz–Makeham formula:

$$\mathrm{GM}_x(r, s) = \sum_{i=0}^{r-1} \alpha_i x^i + \exp\left(\sum_{j=0}^{s-1} \beta_j x^j\right) \qquad r, s \geqslant 0 \tag{3.4}$$

subject to the convention that $r=0$ implies the exponentiated polynomial term only, and $s=0$ implies the polynomial term only. This is non-linear in general, unless either one of r or s is zero, dependent on the nature of the link. The reason for this will become apparent shortly.

3.3 *Link Functions*

Consider the binomial model first. There are a number of possibilities in common use:

(1) The log-odds or logit link function defined by:

$$\eta_x = \log\left(\frac{m_x}{R_x - m_x}\right) = \log\left(\frac{q_x}{1 - q_x}\right)$$

with inverse:

$$q_x = \frac{\exp(\eta_x)}{1 + \exp(\eta_x)}. \tag{3.5}$$

(2) The complementary log–log link function defined by:

$$\eta_x = \log\left(-\log\left(1 - \frac{m_x}{R_x}\right)\right) = \log(-\log(1 - q_x))$$

with inverse:

$$q_x = 1 - \exp(-\exp(\eta_x)).$$

(3) The probit link function defined by:

$$\eta_x = \Phi^{-1}\left(\frac{m_x}{R_x}\right) = \Phi^{-1}(q_x)$$

where Φ is the cumulative distribution function of the standard normal variable, and with inverse:

$$q_x = \Phi(\eta_x).$$

Notice that each link function maps the domain [0, 1] of the probability q_x or, equivalently, the domain [0, R_x] of the mean m_x to the whole of the real line, thereby guaranteeing the basic tenet that the inverse link gives rise to probability values in the interval [0, 1]. It should be noted that use of the inverse odds formula:

$$q_x = \mathrm{LGM}_x(r, s) = \frac{\mathrm{GM}_x(r, s)}{1+\mathrm{GM}_x(r, s)} \tag{3.6}$$

strongly advocated by Forfar *et al.* (1988), in which $\mathrm{GM}_x(r, s)$ is defined by equation 3.4, does not offer this guarantee (unless $r=0$); a possible shortcoming of which the authors are obviously aware. Clearly, when $r=0$, the Forfar *et al.* equation 3.6 coincides with the inverse logit link 3.5 based on a polynomial predictor in x of degree $s-1$.

Consider the Poisson modelling distribution next. It is mainly associated with the log-link function:

$$\eta_x = \log(m_x) = \log(R_x) + \log(\mu_{x+\frac{1}{2}}). \tag{3.7}$$

3.4 *Implementation and Diagnostic Checks*

This is by the GLIM computer software package. A series of macros have been written to facilitate both the fitting and diagnostic checking of individual models as well as the construction of graduated tables. These are available on request. Neither the age scaling transformation nor the orthogonal polynomials discussed by Forfar *et al.* (1988) are needed. However, there would appear to be a limitation on the ability to handle the more general non-linear $\mathrm{GM}_x(r, s)$ predictors defined by equation 3.4 when $s \geqslant 3$. Although designed specifically to handle linear predictors it is nevertheless possible to model with the non-linear $\mathrm{GM}_x(r, 2)$ predictor as described in Section 6.

Graduations are based on models fitted to the ungrouped raw data (A_x, R_x) over a range of xs, in keeping with Forfar *et al.* (1988). The overall measure of goodness of fit is provided by the model deviance, minus twice the log-likelihood ratio statistic based on the current model relative to the so-called saturated or full model. It is possible to compare the goodness of fit of the various modelling predictor structures fitted to the same data set by differencing the resulting model deviances, while no attempt should be made to interpret their absolute values. Asymptotically these differences have the chi-square distribution. Further detail is described in Section 5.

As with Forfar *et al.* (1988), the data are grouped when necessary—so that the expected number of deaths in adjacent age cells exceed 5, say—before residuals are computed. For the binomial models, since the q_x are small (except for very large x), following the first edition of McCullagh & Nelder (1983), the adjusted deviance residuals, defined by:

$$r_x = \pm\,(2A_x\log(A_x/R_x\hat{q}_x) + 2(R_x - A_x)\log((R_x - A_x)/(R_x - R_x\hat{q}_x)))^{\frac{1}{2}} \\ + (2\hat{q}_x - 1)/(6(R_x\hat{q}_x(1-\hat{q}_x))^{\frac{1}{2}})$$

where the sign is that of $A_x - R_x\hat{q}_x$, are preferred. In the case of Poisson models, the deviance residuals:

$$r_x = \pm\,(2(A_x \log(A_x/R_x\hat{\mu}_{x+\frac{1}{2}}) - A_x + R_x\hat{\mu}_{x+\frac{1}{2}}))^{\frac{1}{2}}$$

where the sign is that of $A_x - R_x\hat{\mu}_{x+\frac{1}{2}}$, are computed.

Plots of residuals against both age and against fitted (expected) values are recommended as minimal informal visual checks on various aspects of the modelling assumptions. The residual plot against age augments the more formal tests performed by Forfar *et al.* (1988); a null pattern being consistent with low autocorrelation, an adequate number of 'runs', etc. and consequently potentially sound graduations.

By nominating the binomial or Poisson modelling distributions it follows that the respective variance functions are defined to be:

$$V(m_x) = m_x\left(1 - \frac{m_x}{R_x}\right), \qquad \text{with mean } m_x = R_x q_x$$

and $$V(m_x) = m_x, \qquad \text{with mean } m_x = R_x\mu_{x+\frac{1}{2}}.$$

The residuals plotted against either fitted (expected) values or fitted values transformed to the constant-information scales, defined by $2\arcsin\sqrt{\hat{q}_x}$ or $2\sqrt{R_x\hat{\mu}_{x+\frac{1}{2}}}$ for the binomial and Poisson models respectively, act as an informal check on the adequacy or otherwise of the variance function of the chosen model. The transformation usually has the effect of spreading out the points on the horizontal scale. Again a null pattern showing no trend is consistent with the assumed variance function. These are potentially informative plots to examine, since the presence of duplicate policies on the same lives is known to affect the assumed variance function of the model in the ways outlined by Forfar *et al.* (1988).

4. GRADUATION AND SOME SPECIFIC GLMS

4.1 *The Gompertz GLM*

It is perhaps most fitting to commence a more detailed study of specific models with a GLM motivated by the Gompertz historic and familiar formula in which the force of mortality, μ_x, at age x, is given by the non-linear form:

$$\mu_x = Bc^x.$$

Although μ_x is non-linear, it is well-known—see Chapter 14 of Benjamin & Pollard (1980), for example—that a linearised form is forthcoming in terms of the probability of death, q_x, namely:

$$\eta_x = \log(-\log(1 - q_x)) = \beta_0 + \beta_1 x$$

in which β_0 and β_1 comprise a reparameterisation of B and c. In fact:

$$c = \exp(\beta_1), \qquad B = \log(c)\exp(\beta_0)/(c-1).$$

Thus graduation by Gompertz historic formula can be likened to the fitting of a

GLM comprising the binomial modelling distribution, complementary log–log link, and straight-line predictor ($s=2$). It is proposed to extend the model by allowing the more general polynomial predictor 3.1. Indeed, an examination of Figure 4.1 of Benjamin & Pollard (1980), in which complementary log–logs are plotted (on a negative scale) against age x, suggests that the inclusion of an additional quadratic term in the linear predictor might be an alternative to the use of Makeham's formula. Initial insight into a possible suitable predictor is obtained by plotting the empirical complementary log–logs against age, while the graduated values are determined by:

$$\hat{q}_x = 1 - \exp(-\exp(\hat{\eta}_x))$$

where the estimated linear predictor:

$$\hat{\eta}_x = \sum_{j=0}^{s-1} \hat{\beta}_j x^j.$$

4.2 *The Wilkie GLM*

This implies treating the observed numbers of deaths at ages, x, as independent binomially distributed responses:

$$A_x \sim B(R_x, q_x)$$

in conjunction with the logit link and linear predictor so that the graduated values are determined by:

$$\hat{q}_x = \frac{\exp(\hat{\eta}_x)}{1+\exp(\hat{\eta}_x)}$$

with: $$\hat{\eta}_x = \sum_{j=0}^{s-1} \hat{\beta}_j x^j.$$

The model was first used in this context at the suggestion of Wilkie (CMI Committee, 1976). Indeed all the graduations in the *CMI Report* No. 2 (1976) were constructed in this way.

4.3 *The Probit GLM*

Here this model again implies binomial responses:

$$A_x \sim B(R_x, q_x)$$

independently for each age, x, in conjunction, this time, with the probit link and linear predictor giving rise to the graduated values:

$$\hat{q}_x = \Phi\left(\sum_{j=1}^{s-1} \hat{\beta}_j x^j\right).$$

The model would not appear to have been used in this context before.

4.4 *The Poisson GLM Targeting the Force of Mortality*

This comprises Poisson response variables, A_x, the log-link function with a polynomial predictor written as:

$$\eta_x = \log(R_x) + \log(\mu_{x+\frac{1}{2}}) = \log(R_x) + \sum_{j=0}^{s-1} \hat{\beta}_j x^j$$

which is slighly different in form to the three preceeding cases. Here the $\log(R_x)$ term on the right hand side is treated as an additional regressor variable with a known regression coefficient (value +1). It is necessary to subtract such terms from the responses before fitting the polynomial structure. There is an automatic facility for doing this in the GLIM software package by declaring the $\log(R_x)$ terms as offsets. It follows that:

$$\mu_{x+\frac{1}{2}} = \exp\left(\sum_{j=0}^{s-1} \beta_j x^j\right)$$

and that the model is identical to the so-called μ-graduation, $GM_x(0, s)$ model discussed in Forfar *et al.* (1988). With $s=2$ and a little reparameterisation, Gompertz familiar form is reproduced.

5. STANDARD TABLES AND GLMS

The pensioners' widows graduation (Example 1, Section 15) in Forfar *et al.* (1988) together with the various graduations presented in *CMI Reports* No. 2 (1976) and No. 6 (1983) are based almost exclusively on the Wilkie GLM and provide a most convenient test-bed for comparing the performances of the various binomial GLMs discussed in Section 4. It should, perhaps, be emphasised that these discussions are limited to diagnostic checks based on a comparison of model deviances together with the visual scrutiny of residual plots, and which would appear to be new to existing actuarial practice. It is apparent from the extensive actuarial literature on graduation that many other diagnostics are rightly taken into consideration before a graduation model is finally adopted.

5.1 *The Pensioners' Widows, 1979–82 Experience*

5.1.1 Binomial GLMs with all three of the links discussed in Section 4 and with straight-line and quadratic predictors were fitted to the data presented in Table 15.6 of Forfar *et al.* (1988). The resulting values of the model deviances are presented in Table 5.1.

For a typical specific model structure, the so-called current model, c, the (scaled) deviance is defined by:

$$S(c,f) = -2\log\left(\frac{l_c}{l_f}\right) = 2\log(l_f) - 2\log(l_c)$$

Table 5.1. *Model deviances*

Link	Degree 1	2
1	61·80	61·75
2	61·57	61·53
3	65·19	62·30

(Link: 1-complementary log-log, 2-logit, 3-probit)

where l_c and l_f are the optimum values of the likelihood under the current model and under the saturated or full model, f. The saturated model, f, as beholds this description, is characterised by the property that its fitted values are the empirical responses themselves, thereby ensuring a perfect fit with zero residuals. For binomially distributed responses, $A_x \sim B(R_x,\ q_x)$, as is the case here, with expected values, $m_x = R_x q_x$, and log-likelihood:

$$\log(l) = \sum_x \left(A_x \log\left(\frac{m_x}{R_x}\right) + (R_x - A_x)\log\left(\frac{R_x - m_x}{R_x}\right)\right) + \text{const.}$$

it follows that the model deviance is given by:

$$S(c,f) = 2\sum_x \left(A_x \log\left(\frac{A_x}{\hat{m}_x}\right) + (R_x - A_x)\log\left(\frac{R_x - A_x}{R_x - \hat{m}_x}\right)\right)$$

where $\hat{m}_x = g^{-1}(\hat{\eta}_x)$ denote the fitted values under the current model, c. Then for a given data set, specific link, g, and nested polynomial predictor structures, c_1 and c_2, of degree s_1 and s_2 ($s_1 < s_2$) respectively, the differences:

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

are referred approximately to the chi-square distribution with $(s_2 - s_1)$ degrees of freedom. Further, since the value of l_f is invariant across links for a specific data set, the relative magnitudes of the deviances $S(c,f)$ for a fixed model structure, c, provide a means of comparison across links. There is, however, no reference distribution.

The deviances play the same role as the log-likelihood plus large unspecified constant quoted in Forfar *et al.* (1988). Indeed, the former are clearly a linear (straight-line) transformation of the latter with a scale factor of 2 and in which the value of the large constant is revealed! In particular, since differences in the deviances can be referred approximately to the chi-square distribution, an 'improvement' of at least 4 is sought in the deviance for one extra parameter, corresponding approximately to the 95% point on a chi-square distribution with one degree of freedom. This is, then, equivalent to an improvement of at least 2 in the log-likelihood, an initial sifting criterion used extensively by Forfar *et al.* (1988) to identify possible optimum predictor structures.

Scrutiny of Table 5.1 lends support to the choice of straight-line predictor and marginally supports the choice of logit link over the classical Gompertz formula. The residual plots associated with the fit, some of which are reproduced in Figure 5.1, are highly supportive of the graduations presented by Forfar *et al.* (1988). In

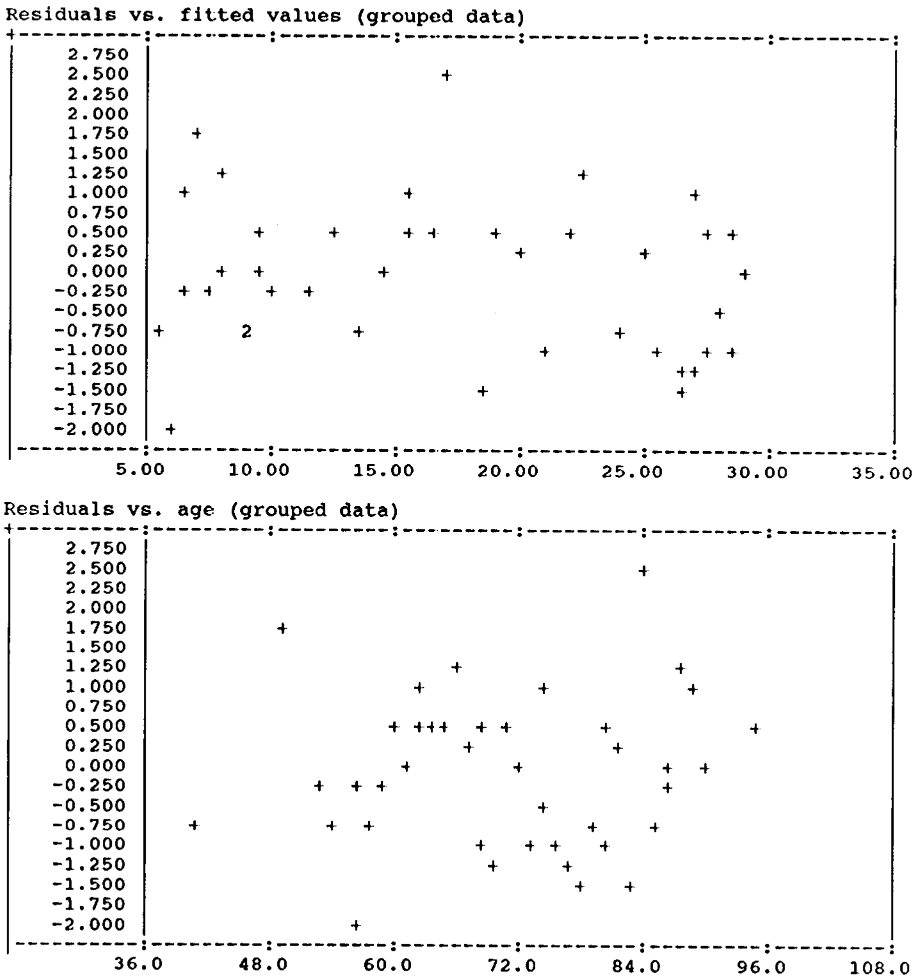

Figure 5.1. Residual Plots: Pensioners' Widows.

particular, the lack of any trend in the residuals plotted against age is consistent with the supportive evidence of the formal tests applied by Forfar *et al.* (1988), while the null pattern in the residuals plotted against fitted (expected) values is supportive of the assumed variance function.

5.1.2 The Poisson GLM with log-link, offsets and with straight-line and quadratic predictors were fitted to the data presented in Table 15.5 of Forfar *et al.* (1988). The resulting deviances are presented in Table 5.2.

Table 5.2. *Model deviances*

Degree	
1	2
60·98	60·94

(Link: log)

For Poisson distribution responses, $A_x \sim P(R_x\mu_{x+\frac{1}{2}})$, with expected values $m_x = R_x\mu_{x+\frac{1}{2}}$ and log-likelihood:

$$\log(l) = \sum_x (A_x\log(m_x) - m_x) + \text{const.}$$

it follows that the model deviance is given by:

$$S(c,f) = 2\sum_x\left(A_x\log\left(\frac{A_x}{\hat{m}_x}\right) - (A_x - \hat{m}_x)\right)$$

where $\hat{m}_x = g^{-1}(\hat{\eta}_x)$ are again the fitted values under the current model.

Table 5.2 lends support to the choice of straight-line predictor, while the residual plots, which are not reproduced here, are highly supportive of the model used by Forfar *et al.* (1988).

5.2 *The* CMI Report *No. 2* (*1976*)

Data are presented for the pensioners' 1967–70 experience for both male and female normal lives and amounts (Tables 3, 4, 5, 6) and for the annuitants' 1967–70 experience for both females and males, each at duration 0 and at duration 1 or more (Tables 14, 15, 19, 20). In the event, with one exception, graduations were based on a straight-line predictor in conjunction with the Wilkie GLM; the exception being that for female annuitants, duration 1 or more, where a cubic predictor was used.

Binomial GLMs with all three of the links discussed in Section 4 and with straight-line, quadratic and cubic predictors were fitted to all eight data sets and the resulting model deviances presented in Table 5.3.

Scrutiny of the deviances in Table 5.3 for all eight data sets separately leads to the following conclusions:

Table 5.3. *Deviances, binomial models, various links, various polynomial predictors*

Pensioners 1967–70: Normal

	Type	Lives			Amounts		
	Degree	1	2	3	1	2	3
	Link						
Male	1	90·27	82·78	82·36	247·40	238·48	238·46
	2	84·96	83·14	82·13	241·21	238·43	238·04
	3	91·71	83·16	82·18	246·35	238·83	237·88
Female	1	53·36	53·28	52·68	89·53	89·01	79·23
	2	53·52	52·87	52·63	89·24	87·33	79·46
	3	64·01	52·85	52·85	102·95	85·38	81·08

Annuitants 1967–70:

	Duration	0			1 or more		
	Degree	1	2	3	1	2	3
	Link						
Male	1	37·78	36·69	34·62	81·15	80·50	80·18
	2	37·92	36·51	34·64	83·01	80·96	80·35
	3	39·78	35·99	34·71	95·00	81·40	80·63
Female	1	52·49	52·14	49·70	74·35	73·19	65·41
	2	53·13	52·60	49·97	75·58	70·96	65·46
	3	56·94	53·41	50·66	113·91	68·30	65·86

(Link: 1-complementary log-log, 2-logit, 3-probit.)

(i) supports the use of a cubic predictor for pensioners, female, amounts as well as for annuitants, female, duration 1 or more, which, presumably, was rejected because of other considerations;

(ii) endorses the use of the linear predictor in the remaining six categories;

(iii) offers strong support for the logit link over the other two links in only two of the eight categories, namely for pensioners, male, both lives and amounts;

(iv) offers evidence, albeit marginal, that Gompertz complementary log–log link is to be preferred over the Wilkie logit link in five out of the remaining six cases, so that the graduations could possibly have been improved upon;

(v) indicates that the probit link performs less well than the other two links in all eight categories; and

(vi) as an example of questionable modelling assumptions, the residuals plotted against both fitted values and against fitted values transformed to the constant-information scale for female annuitants, duration 1 or more, with logit link and cubic predictor are reproduced in Figure 5.2. These possibly cast suspicion on the nature of the variance function. One possible explanation could lie with the presence of duplicate policies in

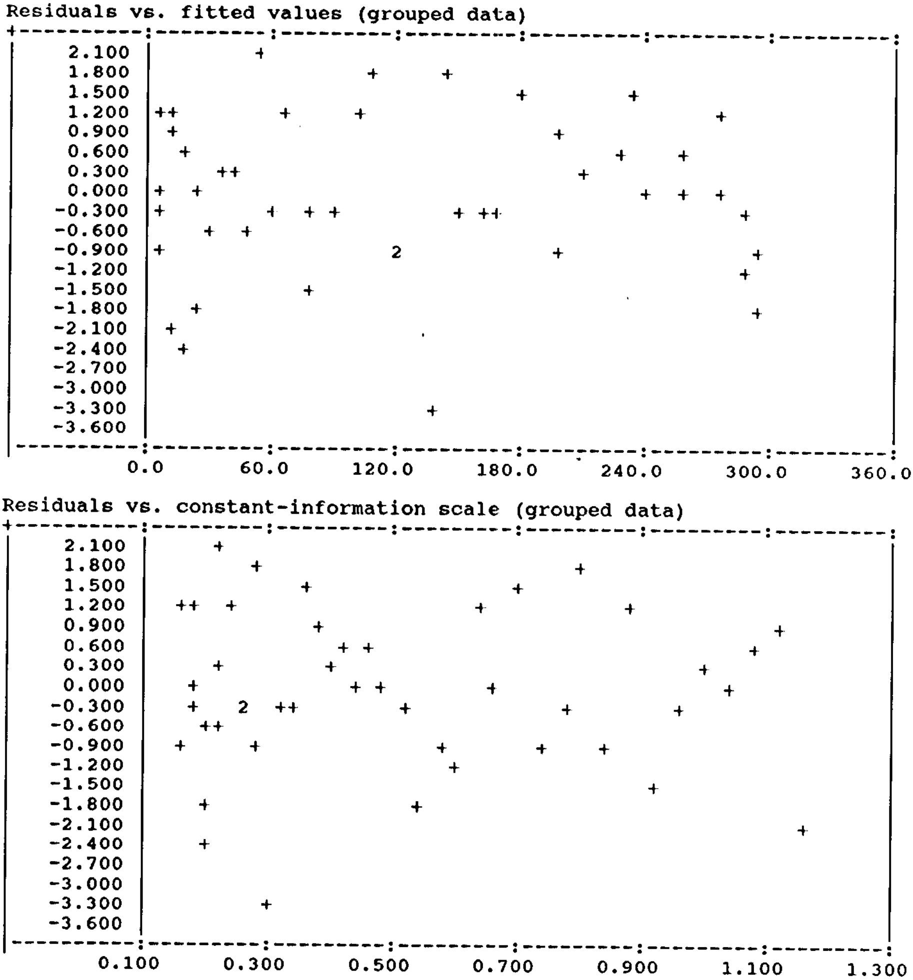

Figure 5.2. Residual Plots: Annuitants, Females, Duration 1+.

the data set. It goes without saying that each of the graduations finally adopted in the *CMI Report* No. 2 were subjected to the battery of formal tests.

5.3 *The* CMI Report *No. 6 (1983)*

Graduations for the assured lives' 1975–78 experience are presented with the ultimate rates of mortality (Appendix 3, pp. 20–21) based on a Wilkie GLM with quadratic predictor. Unfortunately, it has not been possible to reproduce these results exactly, as the relatively small number of experiences in the age range 15–19, used in the construction of the CMI ultimate table, have been omitted from the report. Subject to this very minor omission in the data set, the residual plots are supportive of the model adopted in the Report, but the probit link gives a better fit; a deviance of 86·84 compared with 89·39 for the logit link and 90·26 for the complementary log–log link. Thus, again, it is conceivable that the graduations could have been improved upon using this wider array of modelling techniques. An over parameterised predictor has been selected for the reasons given in the Report.

6. GRADUATION AND SOME SPECIFIC GNLMS

6.1 *The Makeham GNLM*

Again, it is most fitting to commence with a GNLM motivated by an historic formula, namely the three parameter Makeham form:

$$\eta_x = A + Bc^x.$$

It is well known that it is not possible to transform this non-linear form into a linear form unless $A=0$. However, introducing a trivial reparameterisation, it is possible to write:

$$\eta_x = \alpha + \beta \exp(\phi x)$$

which is perceived to be linear if the third parameter ϕ is assumed known. This points the way forward.

In keeping with McCullagh & Nelder (1989) let $g(x;\phi)$, with parameter ϕ, denote a typical non-linear term in any linear predictor. Specifically:

$$g(x;\phi) = \exp(\phi x)$$

here. Expansion, to first order terms, about an initial value ϕ_0 yields:

$$g(x;\phi) \simeq g(x;\phi_0) + (\phi - \phi_0)\left(\frac{\partial g}{\partial \phi}\right)_{\phi=\phi_0}$$

so that, to a first approximation, the non-linear term $\beta g(x;\phi)$ in the linear predictor can be replaced by two linear terms:

$$\beta u + \gamma v$$

with covariates:

$$u = g(x;\phi_0),\ v = \left(\frac{\partial g}{\partial \phi}\right)_{\phi=\phi_0}$$

and where:

$$\gamma = \beta(\phi - \phi_0).$$

So, given a starting value; ϕ_0, covariates:

$$u = \exp(\phi_0 x),\ v = x \cdot \exp(\phi_0 x)$$

are computed, and the parameters β and γ (along with the parameters of any other linear terms present in the predictor) estimated by fitting the model. Then ϕ_0 is updated, using:

$$\phi_1 = \phi_0 + \hat{\gamma}/\hat{\beta}$$

and the process repeated until convergence. Convergence is not guaranteed for distant starting values. A starting value of $\phi_0 = 0{\cdot}0005$ has been found to induce convergence in two or three iterations in all of the many typical data sets graduated in this way. Since μ_x is being targetted for graduation, the Poisson modelling distribution with the natural log-link is selected.

6.2 *The Forfar* et al. *GNLM*

Forfar *et al.* (1988) have proposed a non-linear predictor based on their so-called Gompertz–Makeham formula, $GM_x(r, s)$, defined by equation 3.4. To place these models within the generalised linear and non-linear modelling context it is necessary to consider a number of separate cases.

6.2.1 The graduation of q_x using the binomial modelling distribution in conjunction with the inverse-odds transformation is defined by equation 3.6.

(a) If $r=0$ and $s>0$, then equations 3.4 and 3.6 together yield the Wilkie GLM already discussed in Section 4.

(b) More generally if $r>0$, then equation 3.6 can be trivially rewritten as:

$$q_x = \frac{\eta_x}{1+\eta_x}$$

with inverse:

$$\eta_x = GM_x(r, s) = \frac{q_x}{1-q_x}$$

the odds-link function. It is, therefore, something of a misnomer to refer to equation 3.6 as the logit Gompertz–Makeham formula, since it involves the odds and not the log-odds transformation as this choice of name implies. If, in addition, $s=0$ or 1, the predictor 3.4 in this combination of link is linear, otherwise it is non-linear.

As already intimated, this particular predictor link combination ($r>0$) does not automatically preclude the possible occurrence of negative probabilities. (It may well be a case of the ends justifying the means.) For this reason, the odds-link in combination with the binomial modelling distribution is not automatically available for implementation within the GLIM software package. It can, however, be accommodated through the interactive user own macro facility. There is also the additional complication of the non-linearity of the predictor to be taken into account if $s>1$.

6.2.2 The graduation of μ_x, using the Poisson modelling distribution in conjunction with the identity:

$$\mu_x = \mathrm{GM}_x(r, s).$$

(a) If $r=0$ and $s>0$, then equations 3.4 and 3.7 together yield the Poisson GLM with natural log-link, polynomial predictor and declared offsets $\log(R_x)$ already discussed in Section 4. In addition, this particular model formulation ensures that the force of mortality cannot be negative.
(b) For the more general case, with $r>0$, any s, the modelling technique advocated in Forfar *et al.* (1988) is equivalent to a Poisson GNLM with responses A_x/R_x, weights R_x, identity link and non-linear predictor $\eta_x = \mu_x = \mathrm{GM}_x(r, s)$. This becomes immediately apparent on writing the log-likelihood of this Poisson model:

$$A_x \sim P(R_x \mu_{x+\frac{1}{2}})$$

independently for all x, as:

$$\log(l) = \sum_x R_x\left(\frac{A_x}{R_x}\log(\mu_{x+\frac{1}{2}}) - \mu_{x+\frac{1}{2}}\right) + \text{const.}$$

It should be noted that this particular modelling formulation does not exclude the theoretical possibility of a negative force of mortality.

7. STANDARD TABLES AND GNLMS

Graduations for both male pensioners (Example 2, Section 16) and for male assured lives (Example 3, Section 17) presented by Forfar *et al.* (1988) are based on the non-linear Gompertz–Makeham predictor. We concentrate on the latter example, since the majority of graduations quoted are based on the $\mathrm{GM}_x(2, 2)$ predictor which, although non-linear, can be handled by the GLIM software package as described above.

7.1 *The Male Assured Lives U.K. Experience*

The Poisson GNLM based on the response and weights stated above, with identity link and non-linear predictor $\mathrm{GM}_x(2, 2)$ has been successfully applied to the male assured lives, permanent, U.K. experiences for duration 0 and duration 2–4, confirming the results presented in Tables 17.4 and 17.5 of Forfar *et al.* (1988). The same model has also been successfully applied for duration 5+, confirming the results presented in Table 17.9 of Forfar *et al.* (1988), in which the data were first transformed before modelling by dividing both the exposed to risk and numbers of deaths by the so-called variance ratios for each age derived from the count of duplicates among the deaths from the experience. The residual plots for this model, some of which are reproduced in Figure 7.1, are particularly

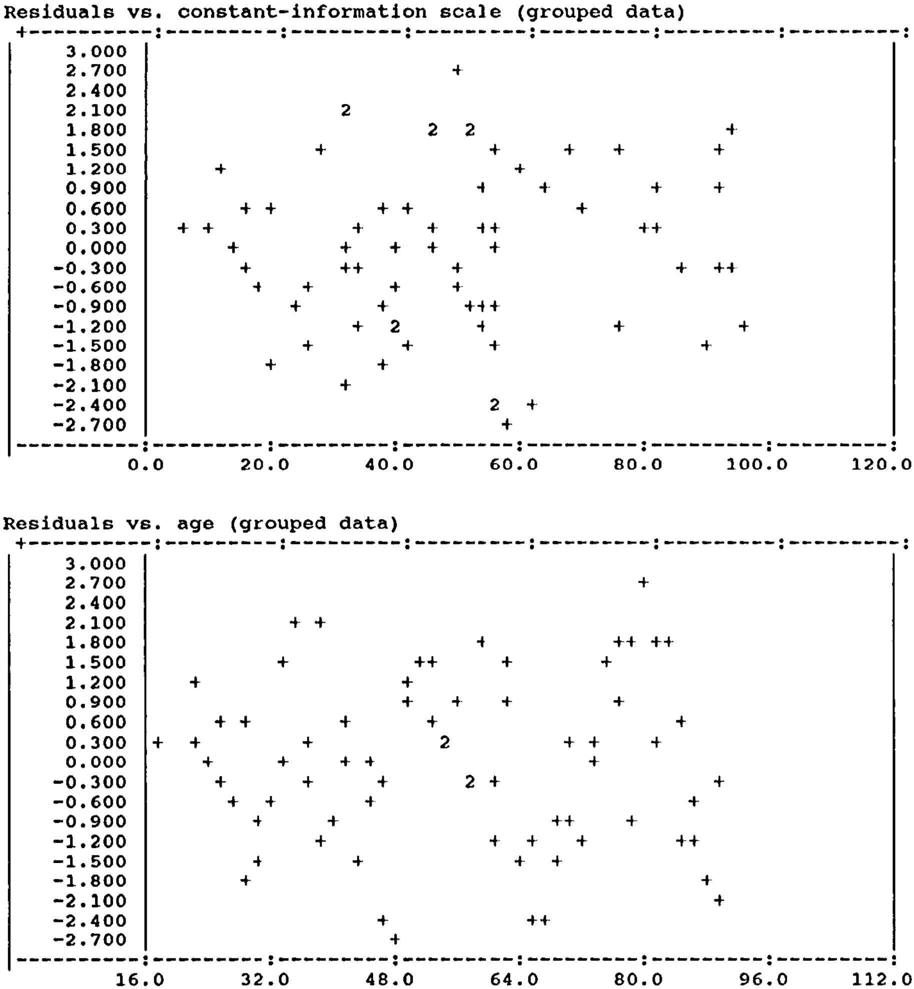

Figure 7.1. Residual Plots: Male assured lives, permanent U.K., Duration 5+.

revealing. The distinctive cyclical pattern in the residuals plotted against age, very marked for ages in excess of 48 years, is consistent with the failure of a number of the formal tests conducted by Forfar *et al.* (1988). However, the null pattern in the residual plot against fitted values transformed to the constant-information scale is supportive of the variance function and is consistent with the data transformation before modelling.

8. SUMMARY

The generalised linear and non-linear modelling frameworks provide the ideal setting for existing actuarial graduation techniques. Not only do these offer a more unified approach to graduation, but they also offer a more comprehensive set of modelling techniques, which have only partially been explored here. In particular, use of parameterised families of link functions and variance modelling techniques in relation to duplicate policies are developed elsewhere. It is particularly gratifying to note how long-established mortality formulae, such as those of Gompertz, in particular, and Makeham, feature within these settings, leading to more general treatments. Further, the graphical analysis of residuals is a highly informative diagnostic check which should not be ignored.

REFERENCES

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

CMI Committee (1976). The Graduation of Pensioners' and of Annuitants' Mortality Experience 1967–70, *CMIR*, **2**, 57.

CMI Committee (1983). Graduation of the Mortality Experience of Female Assured Lives 1975–78, *CMIR*, **6**.

CMI Committee (1988). The Graduation of the 1979–82 Mortality Experiences, *CMIR*, **9**.

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

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