

APPLICATION OF FRAILTY-BASED MORTALITY MODELS USING GENERALIZED LINEAR MODELS

BY

ZOLTAN BUTT AND STEVEN HABERMAN¹

ABSTRACT

Two families of frailty models – Makeham/Gompertz-gamma and Gompertz-inverse Gaussian – have been considered to graduate insurance-based mortality data. The aims of this exercise are twofold. The first aim is to make use of generalized linear models and to evaluate these against traditional techniques. The second aim is to measure the scale of individual heterogeneity in insurance-based populations. The results indicate that (subject to issues of identifiability) there is evidence of frailty in these populations.

KEYWORDS

Frailty, Mortality rates graduation, Generalized linear modelling, Gompertz-Makeham function, Perks model.

1. INTRODUCTION

In mortality analysis, heterogeneity is regarded as any difference in future survival that is potentially observable at the start of a study – for example, at birth or the purchase of an annuity or a life insurance contract. In many cases, heterogeneity may be based on unobserved covariates. Then, the differences in realized survival experience may be attributable to either randomness or unobserved heterogeneity and it may not be possible to discriminate between these two sources. In traditional mortality analysis and insurance applications, heterogeneity has been generally overlooked given that the early survival models tended to provide a good fit for most part of the human lifespan. Departures from observed experience have tended to appear at the oldest ages and have been mainly attributed to scanty data, that is to randomness in the experiment.

Perhaps the first recognition of the effect of heterogeneity is due to Gini (1924) who was commenting on the interpretation of pregnancy rates. The effect in migration was considered by Blumen et al. (1955) and in mortality by the

¹ Faculty of Actuarial Science and Statistics, City University, London.

seminal work of Vaupel et al. (1979) which has led to considerable interest and further developments in the literature. Vaupel et al. introduce the generalized concept of *frailty* represented as an *unobservable univariate* statistical variable Z and describe its implications for standard life table methods (see section 2).

This paper considers practical aspects of selecting the family of frailty models applicable to insurance-based mortality data and the possible interpretation of such models. The paper is organized in six main sections. The next section presents the framework used to formulate the mathematical models for representing frailty. The third section considers the selected models and different methods of fit. In section four we present the insurance data and the graduation results of the formulated models. Then, in section five, there is a short discussion about the interpretation of these results and comparison of our results with other authors' findings. Finally, section six provides some concluding comments.

2. MODEL OF INDIVIDUAL DIFFERENCES IN FRAILITY

As noted by Keyfitz (1985), "... heterogeneity in the underlying population places difficulties in the way of interpretation of all statistical data based on averages". A necessary simplifying assumption in traditional demographic (including life table) analysis has been to include only some of the measurable differences in the composition of a population.

Unlike the effect of sampling variability, which we can deal with by collecting more data, the effect of heterogeneity cannot be reduced by increasing sample sizes. The general consequence of ignoring heterogeneity is that the analysis of population data does not describe the characteristics of any individual but of the cohort as a whole. In mortality applications, the selective effect of heterogeneity is to underestimate the mortality rates at older ages that an average individual will experience. This bias will be explored further in this section.

Following Vaupel et al. (1979), we define *frailty* in a heterogeneous population at any given age x as a random variable Z_x with probability density function $f_x(z)$, such that the individual force of mortality conditional on a realization of $Z_x = z$ will satisfy the following relation:

$$\mu(x|Z_x = z) = \mu_x(z) = z \cdot \mu_x, \quad (1)$$

where μ_x is referred to as the '*standard hazard function*' corresponding to a '*standard individual*', conventionally those with frailty $z = 1$. Here, heterogeneity is modelled in terms of an unobserved scale statistic (a continuous non-negative random variable) that encompasses *all* of the factors affecting human mortality other than age. We note that similar models for the age specific probability of death, q_x , would lead to complexities in terms of the range of values for Z_x , given that q_x is bounded above.

Assuming that there is a unique and fixed value of z associated with a given individual for all his or her life span, Vaupel et al. (1979) note that "... The definition does not imply, however, that individuals at the same level of frailty

are identical – even if they are contemporaries from the same population.” The assumption is merely that their “likelihood of death” would be the same. Ultimately, members of the cohort with higher values of frailty z will have a higher probability of dying and thus the more *frail* individuals are more likely to die first. Such selection effects distort the shape of the population hazard rate, so that this no longer reflects the individual risks.

Based on the proposed model (1) it can be shown (see Butt and Haberman 2002) that the observed population hazard rate is given by:

$$\bar{\mu}(x) = \mu_x \int_0^{\infty} z \cdot f_x(z) dz = \mu_x \cdot \bar{z}_x, \quad (2)$$

where \bar{z}_x is the mean (or expected) frailty among the survivors to age x .

As a consequence of (2), choosing the marginal *p.d.f.* of \mathbf{Z} in the initial birth cohort such that $\bar{z}_0 = 1$ (i.e. ‘standard’ frailty level), the population hazard rate at birth will coincide with the ‘standard’ individual hazard rate $\bar{\mu}(0) = \mu_0$.

More importantly, we can see from the previous expression (2) that, should the expected frailty \bar{z}_x vary significantly with age, the population force of mortality will have a different rate of change with respect to age than the individual force of mortality. Specifically, we note that differentiating \bar{z}_x with respect to age x yields (see Lancaster 1990, page 64):

$$\frac{d}{dx} \bar{z}_x = -\mu_x \cdot \sigma_x^2(z) < 0 \quad (3)$$

where $\sigma_x^2(z)$ is the conditional variance of \mathbf{Z}_x among the population that is alive at age x . Thus, the mean frailty declines with age as death selectively removes those individuals with the higher frailty levels z who are likely to die earlier, and so \bar{z}_x will decrease with age, especially in the case of the older ages. This means that μ_x increases more rapidly than $\bar{\mu}_x$ i.e. “individuals ‘age’ faster than heterogeneous cohorts” (Vaupel and Yashin 1985), so that expectations of life for individuals would be overstated by not allowing for the selective effect of frailty.

The model specified so far contains two critical components:

- a) an age specific ‘standard’ hazard rate for the individuals, μ_x ; and
- b) a choice for the marginal distribution for \mathbf{Z} , $f_0(z)$.²

The choices that have attracted the most attention in the literature are, for a), the Gompertz – Makeham model

$$\mu_x = \alpha + \beta \cdot e^{p \cdot x} \quad (4)$$

and for b) the gamma and inverse Gaussian distributions. Vaupel et al. (1979) make extensive use of the gamma distribution on the grounds that it is tractable

² Given $f_0(z)$, and the form of $\mu_x(z)$, it is possible to derive an expression for $f_x(z)$ for substitution into (2). See Vaupel et al. (1979) and Butt and Haberman (2002) for further details.

analytically, flexible in shape and has positive support (noting that frailty cannot be negative).

In the current paper we consider two underlying frailty distributions, that is the gamma and inverse Gaussian, while the individual hazard is assumed to follow a Gompertz type model. For a further mathematical discussion of these choices see Butt and Haberman (2002).

3. MODELS AND METHODS OF FIT

3.1. Introduction

We consider two methodologies for graduating parametric frailty models: the methods of (unweighted and weighted) non-linear least squares (NLM) and of generalized linear modelling (GLM).

Preliminary results from the NLM approach (weighted and unweighted) have been obtained and compared with those arising from the GLM approach. However, we focus on the GLM results. We have also conducted a systematic sensitivity analysis by allowing for changes to the age range being fitted and the removal of outliers. Finally, trends in the estimated parameters for the population hazard function, and indirectly for the chosen frailty distribution, have been also investigated. The proposed parametric models under investigation are:

a) Gompertz-gamma, known as Perks (P)

$$\bar{\mu}(x) = \frac{a}{1 + e^{b-p \cdot x'}}; \quad (5)$$

b) Gompertz/Makeham-gamma, known as Modified Perks (MP)

$$\bar{\mu}(x) = \alpha + \frac{a}{1 + e^{b-p \cdot x'}}; \quad (6)$$

c) Gompertz-inverse Gaussian (GI)

$$\bar{\mu}(x) = \frac{e^{-d+p \cdot x'}}{\sqrt{1 + e^{-b+p \cdot x'}}}. \quad (7)$$

where $x' = x - 40$ is the age variable adjusted with respect to the minimum age in the observations, in order to improve the parameterization of the models. Models (5) and (7) are presented in outline form in the Appendix. A detailed specification is given in Butt and Haberman (2002).

3.2. Exploratory Data Analysis

It is normal practice to undertake a visual inspection of mortality rates plotted against age before attempting any graduation exercises. As noted by Horiuchi and

Coale (1990) and others, it can be difficult to distinguish, from plots of the logarithms of mortality rates, the differences between a straight line and, say, an upwardly or downwardly concave curve. Horiuchi and Coale propose a different type of exploratory check, by considering the behaviour of

$$\lambda(x) = \frac{d}{dx} \ln \bar{\mu}(x) \quad (8)$$

for different curves. It is clear that if $\bar{\mu}(x) = \beta \cdot e^{p \cdot x}$ so that there is no heterogeneity present then $\lambda(x)$ would be constant. However, for the Makeham-gamma frailty model (6) it is possible to demonstrate that, if

$$\begin{cases} \alpha > 0, & \lambda(x) \text{ is bell shaped with a maximum at age } x_0 \\ \alpha = 0, & \lambda(x) \text{ is monotonically decreasing.} \end{cases}$$

(Indeed we can show that $x_0 - 40 = \frac{b}{p} + \frac{1}{2p} \ln\left(\frac{\alpha}{\alpha+a}\right)$).

Similarly, for the Gompertz-inverse Gaussian model (7), it is straightforward to show that $\lambda(x)$ is monotonically decreasing.

Using some crude estimates for $\lambda(x)$, based on finite difference methods, for a sample of the data sets under consideration, we have noted that there is very little evidence of a bell-shaped profile, indicating the likelihood that $\alpha = 0$. The profiles also tend to be decreasing functions of age, indicating the presence of heterogeneity and supporting the choice of (5) or (7).

An equivalent test would be to consider a plot against age of

$$\lambda(x) = \ln[\bar{\mu}(x+1)^{-1} - \bar{\mu}(x)^{-1}] \quad (9)$$

which, for the Gompertz-gamma model, leads to a downward sloping straight line.

3.3. Generalized Linear Models with Parameterized Link Function

The family of Generalized Linear Models (GLM) has been developed to apply a wide range of (usually) linear regression functions with an error structure which is different from the classical assumptions of a constant variance and a normal distribution. The underlying principle is a natural extension of the classical linear model, with the independent variables present only through a linear configuration. The GLM method allows us to select the appropriate distribution of errors from a comprehensive class of density functions belonging to the exponential family and to derive maximum likelihood estimators. McCullagh and Nelder (1989) provides a detailed review of the methodology whereas actuarial applications of GLMs have been developed by Renshaw (1991) and reviewed by Haberman and Renshaw (1996).

In the current approach, the force of mortality is selected as the target random variable, and we use the overdispersed Poisson distribution to model the numbers of deaths A_x . We also make the simplifying assumption that the central exposures are constant (non-random) quantities for each observational

unit. Considering the responses to be $\hat{\mu}_x = A_x/r_x$, the mean and variance can be expressed as:

$$E\left[\frac{A_x}{r_x}\right] = \frac{E[A_x]}{r_x} = \frac{m_x}{r_x} = u_x, \tag{10}$$

$$\text{Var}\left[\frac{A_x}{r_x}\right] = \frac{\text{Var}[A_x]}{r_x^2} = \frac{\varphi \cdot m_x}{r_x^2} = \frac{\varphi}{r_x} u_x, \tag{11}$$

where u_x is defined to be m_x/r_x . In order to satisfy the overdispersed Poisson distribution assumptions, it is necessary to set the GLM weights equal to r_x .

Observe that the Perks frailty model (5) can be rewritten in terms of the mean value as:

$$u_x = \frac{a}{1 + e^{\eta_x}}, \tag{12}$$

where $\eta_x = b - p \cdot x'$. Although it is not possible to form a linear predictor that would include all three parameters a , b and p , we can still apply the GLM methodology, if we are able to find a continuous and differentiable link function $g(u_x, a) = \eta_x$, with fixed parameter a . Then, the graduation could be performed over a chosen range of values for parameter a assuming that the resulting 'deviance profile' $D(\bar{\mu}; \hat{u}_a)$ is a smooth function with preferably one (global) minimum value D_{\min} over the selected interval. Then the final estimate \hat{a} will be given by the value that minimizes $D(\cdot)$.

Deviance Profiles / Females Duration 5+

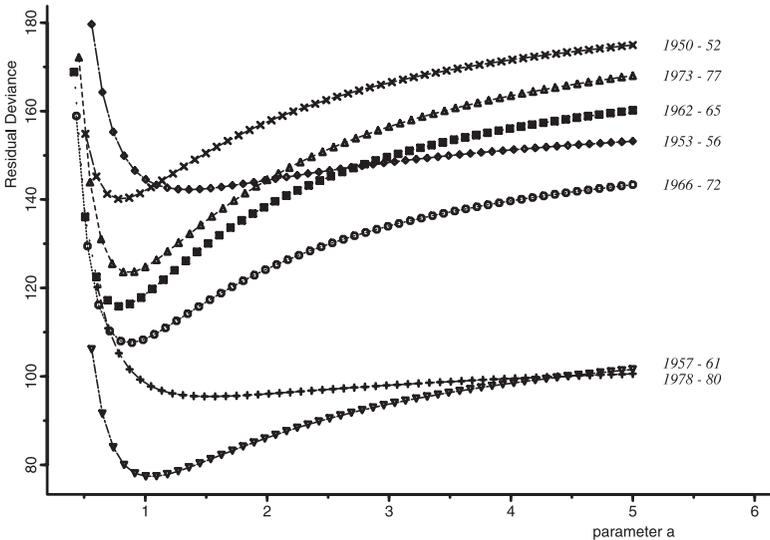


FIGURE 1: Deviance $D(\bar{\mu}; \hat{u}_a)$ vs. fixed values of parameter a for Female Annuitants.

Thus, we can modify the original expression (12) by isolating $\exp(\eta_x)$ to give the following 'parameterized link' function:

$$g(u_x, a) = \ln\left(\frac{a - u_x}{u_x}\right) = \eta_x. \quad (13)$$

This approach has been used by Renshaw (1995). The value of the parameter a has to be pre-determined before the GLM can be fitted. For the structure under consideration, it is possible to search for the optimum value of a by refitting the same predictor structure η_x for different values of a , chosen carefully so that a deviance profile can be constructed.

Note that there is a lower bound imposed on the values which the parameter a can take, since the argument of the logarithm must be positive, that is $a > \max(u_x)$ (see also the interpretation of a in section 5.2). Although we do not prove here that the resulting deviance function $D(\bar{\mu}; \hat{u}_a)$ is continuous and has a single minimum over the domain $(a_{\min}, \infty]$, the results from extensive trials on a range of data sets indicate that this is the case. Figure 1 provides examples of deviance profiles based on the Female Annuitants' Ultimate (i.e. policy duration over 5 years) experience for a range of 8 calendar periods.

4. APPLICATION TO INSURANCE DATA

The proposed models have been applied to a series of large insurance data sets covering an extended historical period, that was collected by the CMI Bureau. For reasons of space, we report here results for two of these data sets, which concern respectively the mortality experience of immediate annuitants and assured lives. Table 1 provides some summary information on these data sets.

TABLE 1
EXTENT OF THE ANALYZED DATA SETS (MILLIONS)

Data Set		Calendar Years	Total		Total	
			Number of Deaths		Exposed to Risk	
			Policy Duration (yrs)		Policy Duration (yrs)	
			1-4	5+	1-4	5+
Immediate Annuitants	M	1946-1994	0.048	0.040	0.599	0.429
	F	1946-1994	0.130	0.116	1.938	1.547
Assured Lives	M	1924-1994	0.071	1.035	9.215	68.960

The data comprise tabulations of observed number of deaths a_x and initial exposures r_x^i by age x and policy duration d . The initial exposure counts have been transformed to central exposures by the approximate formula $r_x = r_x^i - \frac{1}{2} a_x$ to estimate the population hazard rates. Also, grouping by calendar years (and some ages) have been applied in order to avoid the occurrence of zero mortality rates in individual cells and thus *infinite* weights in the WNLM graduation method.

4.1. Immediate Annuitants Data Set

The data consist of cohorts of males and females with lives born in the period between 1846-1943 and followed up between 1946-1994 (note that the observations for years 1968, 1971 and 1975 are not available). The data have been originally recorded by grouped ages 40-45, 45-50, 100+ and for individual ages between 50-100, and also by duration periods (1-4 years and 5+, subsequently rearranged as 1+, i.e. 1-4 and 5+, i.e. 5 and over).

We have grouped the calendar years so that the central year of any group does not coincide with the end of a decade (to avoid bias):

1946-49 (4)	1950-52 (3)	1953-56 (4)	1957-61 (5)
1962-65 (4)	1966-72 (7 - 2 = 5)	1973-77 (5 - 1 = 4)	1978-80 (3)
1981-85 (5)	1986-90 (5)	1991-94 (4)	

where the numbers in parentheses are the number of participating observation years. Graduation has been carried out principally on individual ages between 50 and 100+ with minor changes from one method to another. We note that, for immediate annuitants, the female data set is larger than the male data set and we focus here on results for the female data.

4.2. Annuitants Experience

In the first stage, all the models presented have been fitted to the annuitants' experience using all the available data points with both graduating techniques (NLM and GLM), and consideration has been given not only to the final parameter estimates, the goodness of fit tests³ and examination of the residual plots, but also to the sensitivity of the results to changes in the starting conditions.

Comparing initially the results for the WNLM regressions, it is concluded that it would not be possible to distinguish between the models based on the goodness of fit tests alone (see Butt and Haberman 2002). However, from the detailed parameter estimates and their standard errors, we find that the $\hat{\alpha}$ estimates of the MP model are not significantly different from zero and in some cases are negative. This result is not surprising since our data set focuses on ages over 50, thereby excluding younger persons. Similarly, the \hat{a} parameter estimates for the P and MP models are, in some cases, not significant. On the other hand, all of the parameter estimates for the GI model are significant.

Therefore, based only on this criterion, one might select the GI as the best model for this data set, especially when considering the intuitive property of a non-constant coefficient of variation for the inverse Gaussian (see Butt and Haberman 2002). However, it is generally accepted that the population hazard is best described by a sigmoidal shape curve (i.e. logistic family function), which

³ Appropriate diagnostic tests have been applied to all of the models investigated, based on a visual inspection of plots of Pearson residuals (see Renshaw 1991, 1995) and computation of the traditional graduation tests (i.e. signs, runs and serial correlation tests).

Comparison of the Parameter Estimates resulted from the main models / Females Duration 1+

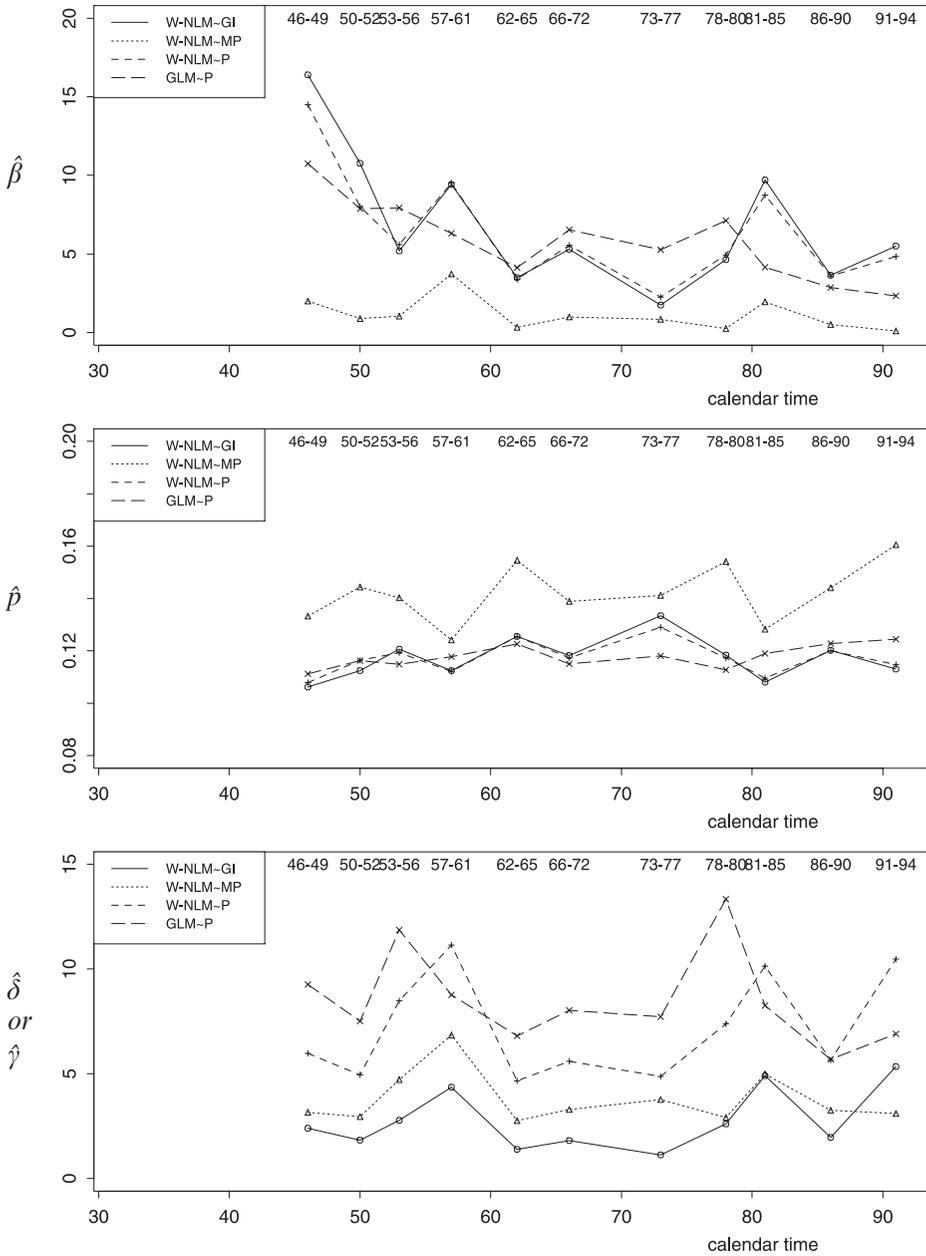


FIGURE 2: Gompertz and Frailty distribution parameters for Female Annuitants Duration 1+ (age range of 40-100+).

Comparison of the Parameter Estimates resulted from the main models / Females Duration 5+

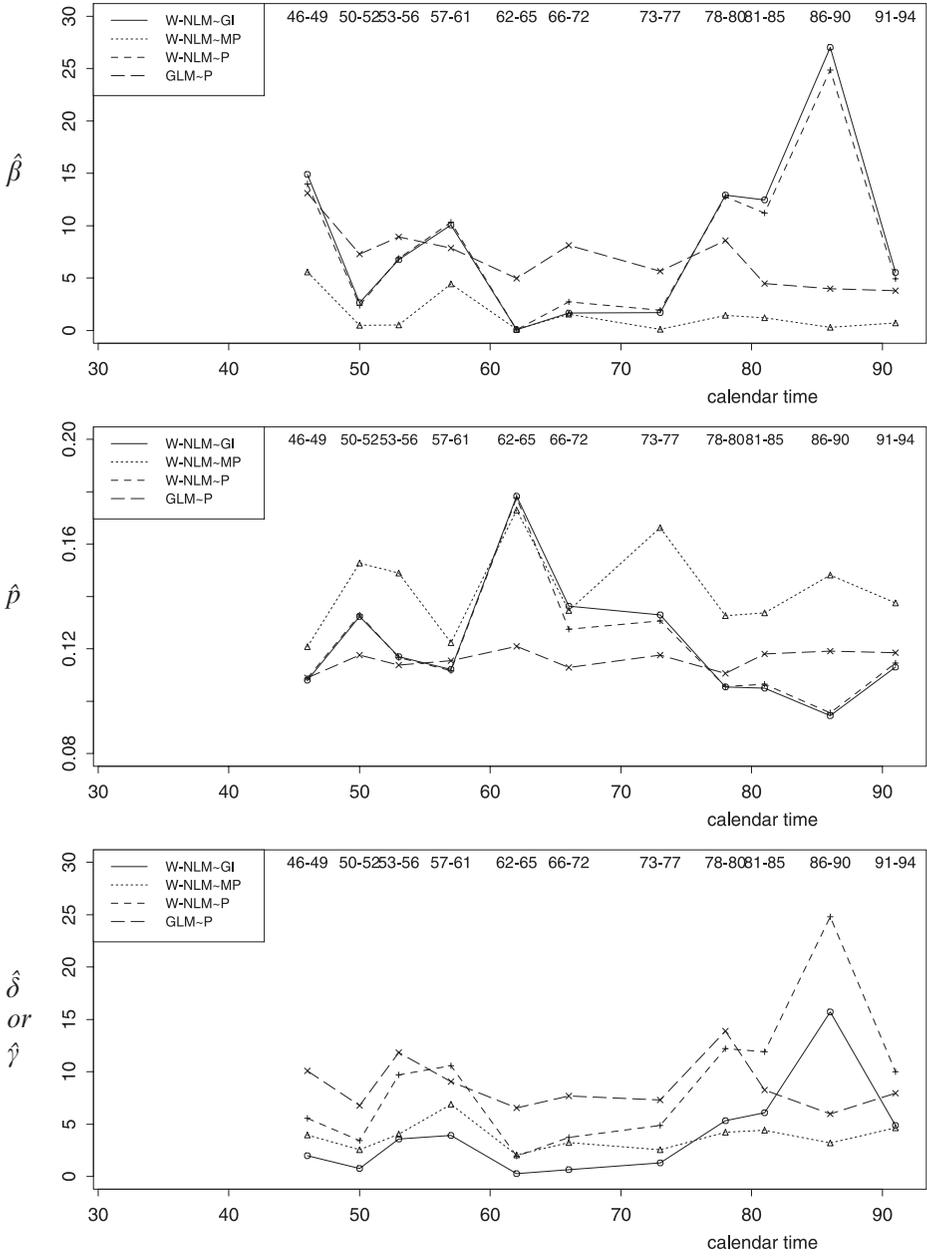


FIGURE 3: Gompertz and Frailty distribution parameters for Female Annuitants Duration 5+ (age range of 40-100+)

for older ages tends to a positive finite value (Thatcher 1999). The GI model does not have this characteristic and it reduces to an exponential function as $x \rightarrow \infty$ (i.e. Gompertz family).

Detailed examination of the residual plots shows that, in general, the P model produces a better fit to the data than the GI model. Therefore, we have opted, on balance, for the P model while keeping in mind that our particular aim is then to use these results as a starting point for taking advantage of the improved regression properties of the GLM method.

We can obtain a good general impression of the features of the models by looking at the estimated parameters of the individual hazard function, $\hat{\beta}$ and $\hat{\rho}$. Similarly we consider the estimates of the shape parameters of the assumed frailty distributions, $\hat{\delta}$ and $\hat{\gamma}$ for gamma and inverse Gaussian respectively, based on the original assumption of $\bar{z}_0 = 1$. The estimators are functions of the graduated population hazard rate parameters (for further details see Butt and Haberman 2002). Further, we note that, for all three frailty models parameter p is common to both the individual and population mortality rate functions.

In Figures 2 and 3, we present the parameter estimates arising from the WNLM and GLM approaches for the female annuitants data set with two duration periods (1+ and 5+) and 11 grouped observation periods. We can see from Figure 2 that the estimates $\hat{\beta}$ and $\hat{\rho}$ of the GI and P models have an almost identical pattern over time, confirming the similarity of these two models for the given age range (as previously suggested based on the goodness of fit tests), whereas the MP model differs significantly from these. It is interesting to note that the estimates based on the GLM method for the P model are within the same range as those given by the GI and P models fitted by the NLM method, but display a much smoother variation by period. Also, in the case of the frailty distribution parameter ($\hat{\theta}$ or $\hat{\gamma}$), all of the models lead to the same type of pattern but at different levels. In the case of durations 5+, Figure 3 shows that the GLM-P model results in more stable estimates.

We base our final model assessment on the following criteria:

- a) the model allows for inhomogeneity in the variance of the errors; and
- b) the parameter estimates are not influenced by the outliers in the data.

The better performance of the GLM regression method could be associated with the fact that it is less influenced by the outliers of a given data set and allows for non-constant variance of the errors. While the WNLM method also caters for the inhomogeneity in the variance of the errors, we note that it is strongly influenced by the outliers. For example, looking at the WNLM graduation results for 1953-56 males' experience in Figure 4, we can see that in the case of W-P and W-MP the fitted curves are pulled down at the oldest ages by the presence of a single outlier. The NLM method is as equally unaffected by the outliers as the GLM; however it fails at the first criterion. In conclusion, the GLM graduation method applied to the P model performs reasonably well under both criteria.

Taking the above considerations into account, we conclude that the GLM method has a stronger theoretical justification and yields models with more

NLM Sensitivity to Outliers / Males Duration 5+

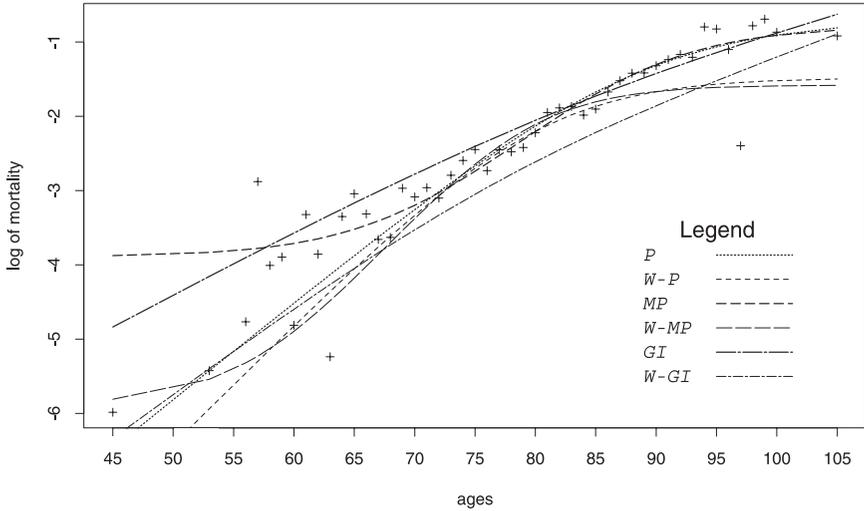


FIGURE 4: NLM Fitting of Perks, Modified Perks and Gompertz-inverse Gaussian models for Male Annuitants Duration 5+ 1953-56.

favourable properties than the classical non-linear least squares method. However, some difficulties have been experienced in relation to the non-linear parameters in the model (e.g. a). Thus, in applications of the GLM-P model, in some isolated cases, we find that:

- i) the minimum value of the deviance curve (see for example, case 1 in Figure 5) lies below the lower bound for parameter a – see section 3.3; or
- ii) the deviance curve tends asymptotically to a minimum only when a increases from the lower bound to infinity (see cases 2-4, Figure 5), instead of having a single global minimum for finite values of a (as in all the cases of Figure 1); and we note that
- iii) there is no straightforward method of estimating the standard error for the parameter a .

In relation to cases i) and ii) above, we note that Butt and Haberman (2002) describe an alternative version of the GLM approach (based on a linearization method and Taylor expansion, advocated by Renshaw 1991) which would provide guidance to the estimate of \hat{a} .

We next examine the sensitivity of the model to the choice of age range and to the influence of outliers. We make use of the chosen model and graduation method, in the form of GLM-P. We restrict the age ranges first to 55-100+ and then to 60-100+ and then investigate the results from fitting the model to the latter age range, with outliers excluded, i.e. ages 60-100. Since a number of arbitrary factors are involved, this part of the analysis is more informal in

Unsuccessful Deviance Profiles / Male Annuitants

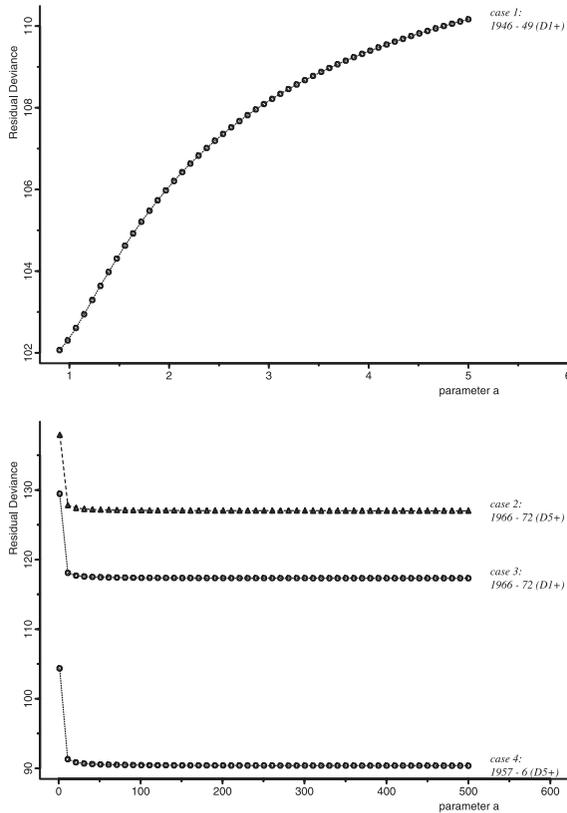


FIGURE 5: Difficulty with the non-linear parameter in (GLM) fitting, either $\min(a) < \max(\mu_x)$ or $Dev \rightarrow \min$ only if $a \rightarrow \infty$.

nature. Evidently, there are some significant effects resulting from the elimination of outliers, as illustrated in the next paragraph; however, these effects are less dramatic than those offered by the NLM or WNLM approaches in relation to the same conditions.

In the case of the annuitants' experience, restricting the age range means, in effect, the elimination of most of the 'noise' at the extremes (in the case of the annuitants' experience, this corresponds to observations which are mainly available only for grouped ages, like 40-45, 45-50 or 100+). This change has little or no effect for the female annuitants' data in terms of the fitted values, the goodness of fit tests, and the frailty distribution parameter $\hat{\delta}$. This can be observed when comparing the corresponding curves from Figures 2 and 6. However, regarding the sensitivity to the outliers, we can note that there are, not unexpectedly, some changes in the model parameter estimates when the outliers are removed. Although these have very little effect on the predicted values – and we observe improvements in the goodness of fit tests – we note that

Sensitivity of GLM-P parameter estimates to outliers Female Annuitants Duration 1+

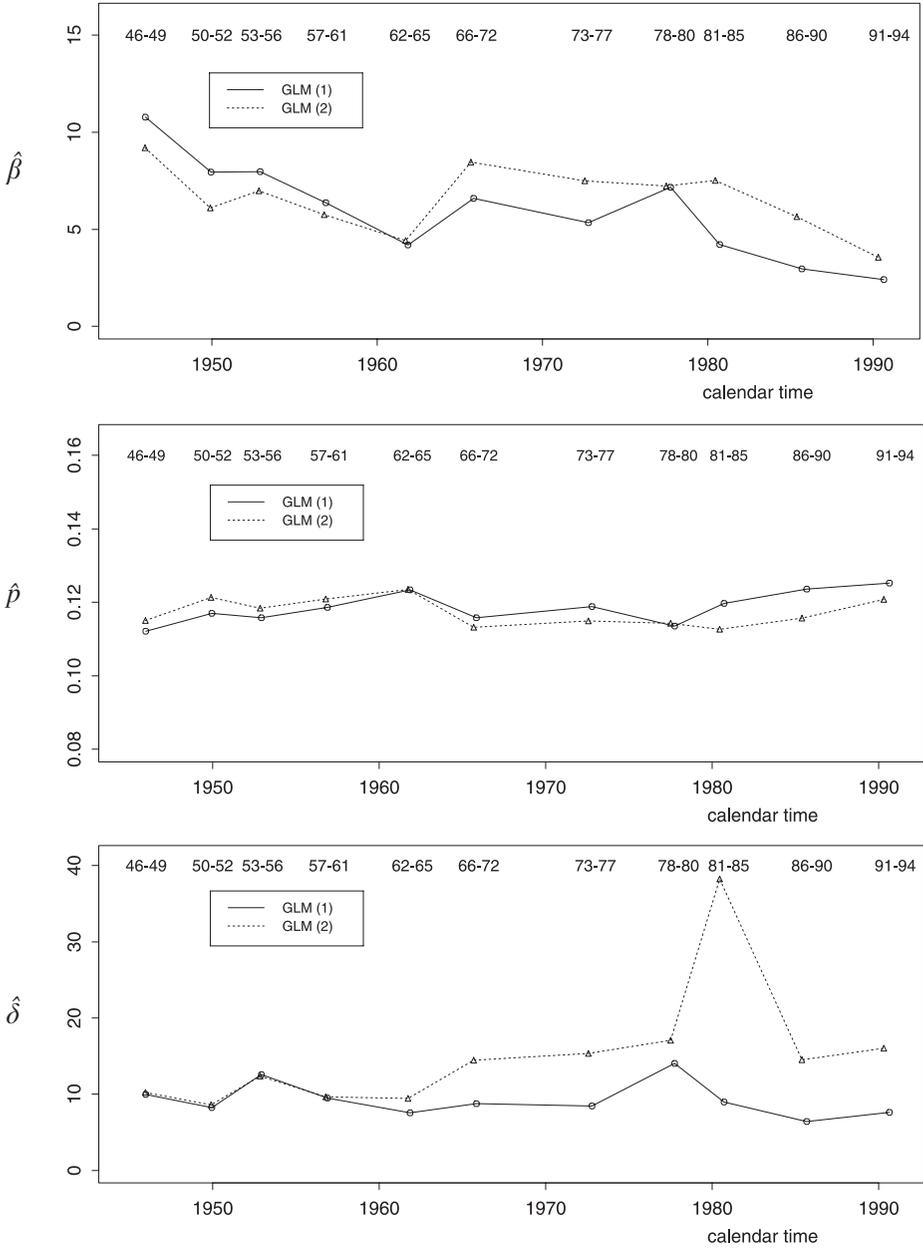


FIGURE 6: Results of GLM-P model fitted to age range 60-100 applied to data sets with (1) and without (2) outliers.

the effect is concentrated on the $\hat{\beta}$ and $\hat{\delta}$ estimates, as can be seen in Figure 6. We also note an upward trend over time for the frailty shape parameter $\hat{\delta}$ for duration period 1+ (and 5+: not shown here).

Butt and Haberman (2002) provide detailed parameter estimates of the above graduations.

4.3. Assured Lives (Males) Data Set

This is the largest data set analyzed and it is in a similar format to the annuitants' data set. The experience of male policyholders, however, dominates and is the focus of our analysis. The recorded age range is considerably longer than the annuitants' data set – generally individual ages between 10-100+. The available duration periods have been used in the form 0, 1, 2-4 and 5+, in line with other authors. We note that the data in the first three groups, corresponding to lower durations, are significantly smaller in size and less reliable than the 5+ group. The data are available for groups of 4 years, with the exception of the last 4 calendar years 1991-94, where the data for individual calendar years are also given:

1924-28 (5)	1929-33 (5)	1934-38 (5)	1949-52 (4)
1953-58 (6)	1959-62 (4)	1963-66 (4)	1967-70 (4)
1971-74 (4)	1975-78 (4)	1979-82 (4)	1983-86 (4)
1987-90 (4)	1991-94 (4)		

As before, the numbers in parentheses are the number of contributing years.

4.4. Assured Lives Experience

In the light of the discussion in section 4.2, we focus here on the application of the Perks (P) model and GLM method to the assured lives data set. These data are particularly extensive for ultimate durations 5+ and are of good quality. Since the graduation results for the lower durations (1, 2-4) turn out to be inconclusive (for all the methods and models), we present the results here only for the ultimate durations 5+. We also note that we have limited the graduation to age ranges 50-100+ in order to allow comparison of results with those obtained in section 4.2 (although it could also be worthwhile to examine the parameter $\hat{\alpha}$ when the MP model is fitted over the full age range).

When applied to the initial age range of 50-100+, the graduations are generally stable and yield parameters which are significantly different from zero throughout the whole data set. It is important to note, however, that the goodness of fit tests do not support all of the fitted models. Thus, the runs and serial correlation tests (which are not independent) show weak graduations for some cases, for example for the periods 1953-58, 1987-90 and 1991-94. The tests indicate strong dependence between individual observations and this is also highlighted by the diagnostic plots which suggest cyclic patterns in the residuals. Further we note that, for the last two groups of observations, the estimates

Sensitivity of GLM-P parameter estimates to outliers Male Assured Lives Duration 5+

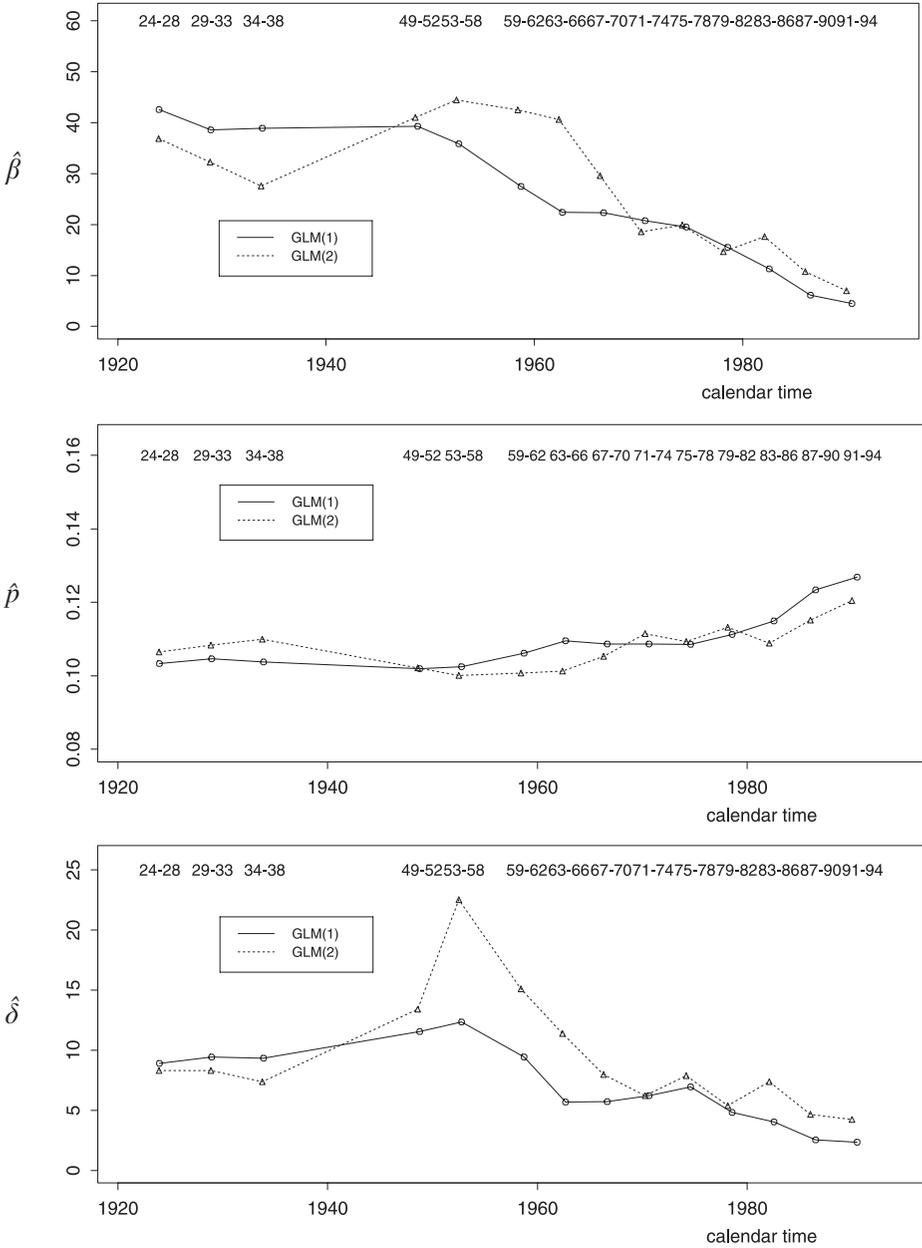


FIGURE 7: Results of GLM-P model fitted to age range 60-100 applied to data sets with (1) and without (2) outliers.

of \hat{a} are very low, 0.253 and 0.237 respectively, suggesting that the models are unsuitable in the light of the interpretation of this parameter as the upper bound for the force of mortality (see section 5.2).

However, when we successively restrict the age ranges (to 55-100+ and 60-100), automatically removing some of the ‘noise’ from the data at the very old ages, the \hat{a} estimates generally improve together with the goodness of fit tests, except for the 1987-90 data which yield a poor runs test. Then, when we remove all outliers and we graduate the age range 60-100, we note an overall improvement for most calendar year groups. The exceptions are again the last two periods mentioned above, where although the estimates of \hat{a} increase to 0.424 and 0.391 respectively, the graduations still produce poor fits, judged by the goodness of fit tests.

Figure 7 presents the trends in the parameter estimates arising from the above sensitivity tests for the male assured lives. It is interesting to note that the frailty parameter estimates, $\hat{\delta}$, are on average lower than for the female annuitants data sets (see Figures 2 and 3), and the profile as a function of time is peaked with a maximum in the mid 1950’s and a subsequent decline.

Butt and Haberman (2002) provide detailed parameter estimates of the above graduation.

5. INTERPRETATION OF RESULTS

5.1. Identifiability

There are difficulties with interpreting the results of the model fitting exercise described in previous sections and interpreting the suitability of, say, the Perks model for describing mortality variation with age.

There is an identifiability problem. As noted by Hougaard (1984): “The frailty distribution is not identifiable, if the frailty is an individual quantity. ... It is not possible to divide the variation into that within and that between individuals, if there is only one observation per individual.”

Yashin et al. (1994) demonstrate that mortality data on their own cannot definitively be used to identify the underlying mechanism of survival. They show that, without additional covariates or assumptions, the fixed frailty Gompertz/Makeham-gamma model cannot be distinguished from a range of models incorporating changing frailty at the individual level (for example, Le Bras 1976).

This conclusion follows on from Hoem (1990) who proves that many underlying mortality rates and many frailty distributions produce the same mortality pattern. As a specific example, we consider the Gompertz-gamma model:

$$\bar{\mu}(x) = \frac{\delta}{\theta + H_x} \mu_x \quad (14)$$

where $\mu_x = \beta \cdot e^{p \cdot x}$ and $H_x = \int_0^x \mu_t dt$.

Then, given a non-negative random variable with distribution function $G(t)$ and finite mean, it is possible to construct a new frailty model with individual

force of mortality (corresponding to $Z = 1$) v_x and frailty distributed at birth with $f_0(z) = G'(z)$ such that the population hazard rate matches (14) exactly. The form of v_x is:

$$v_x = \frac{\delta \cdot \mu_x}{\theta + H_x} b\left(\delta \ln\left(1 + \frac{H_x}{\theta}\right)\right) \quad (15)$$

where $b(\cdot)$ is a function defined by the following steps:

$$r(x) = \int_0^\infty u \cdot e^{-u \cdot x} dG(u) \Big/ \int_0^\infty e^{-u \cdot x} dG(u)$$

$$R(x) = \int_0^x r(y) dy \quad \text{and} \quad b(x) = \frac{d}{dx} R^{-1}(x).$$

5.2. Interpretation of the Perks Model Parameters

The parameters for the Perks model (5) have a useful interpretation, as follows:

1. a is clearly the limiting asymptotic value for the population force of mortality:

$$a = \lim_{x \rightarrow \infty} \bar{\mu}(x).$$

2. In the Gompertz model ($\mu_x = \beta \cdot e^{p \cdot x}$) it is clear that p measures the relative rate of increase in the force of mortality with age. For the Perks model, we note that, following Thatcher (1999):

$$\lambda(x) = \frac{1}{\bar{\mu}(x)} \frac{d}{dx} \bar{\mu}(x) = p \left(1 - \frac{\bar{\mu}(x)}{a}\right). \quad (16)$$

Thus, the relative rate of increase is close to p at young ages, where $\bar{\mu}(x)$ is small, but then it reduces with age. For example at ages high enough for $\bar{\mu}(x)$ to be approximately $\frac{1}{2}a$, the relative rate of increase would have fallen to $\frac{1}{2}p$.

3. b can be interpreted via the age $x_1 = 40 + \frac{b}{p}$, that is the age at which $\bar{\mu}(x) = \frac{1}{2}a$, i.e. half its asymptotic value, and the age at which the Perks curve has its point of inflection.

5.3. Comparison of Results

It will be useful to compare the result with those obtained from other similar studies.

Estimates, \hat{p} , are available from a number of recent studies. Thus, the full Heligman-Pollard model has been fitted by Forfar and Smith (1988) to successive English Life Tables. At the oldest ages, this model reduces to

$$q_x = \frac{G \cdot H^x}{1 + G \cdot H^x} \quad (17)$$

and so $\ln H$ is directly comparable to \hat{p} in our notation, setting aside the difference between q_x and μ_x . For ELTs No 11 (1950-52), 12 (1960-62) and 13 (1970-72) the estimates for \hat{p} would be 0.10003, 0.10346 and 0.10373 respectively. However, we note that this exercise involved fitting to the already graduated q values rather than to the crude observations.

Congdon (1994) uses adult mortality data (ages 25-90) for Greater London for 1980-82 and fits a range of models, including

- a) one based on the ELT No 12 graduation formula and
- b) the modified Heligman-Pollard model

$$q_x = \frac{G \cdot H^x}{1 + \kappa \cdot G \cdot H^x} \quad (18)$$

The estimates corresponding to \hat{p} are

- a) 0.0716 and 0.1086 for males and females respectively
- b) 0.10400 and 0.09685 for males and females respectively.

We note that in the case of b) the estimate, $\hat{\kappa}$, is negative for females: this would imply that $q_x \rightarrow \infty$ as $x \rightarrow \bar{x}$, where $1 + \kappa \cdot G \cdot H^{\bar{x}} = 0$.

Wang and Brown (1998) fit the Gompertz-gamma model to the 1994 Group Annuity Reserving Table of the Society of Actuaries, obtaining the following estimates, in our notation:

$$\hat{a} = 0.15359 \quad \hat{b} = 5.20153 \quad \hat{p} = 0.11761 \quad \hat{\beta} = 7.762 \times 10^{-6} \quad \hat{\theta} = 1.306$$

Given the interpretation of a as $\lim_{x \rightarrow \infty} \bar{\mu}(x)$, this particular estimate \hat{a} looks very low!

Manton et al. (1986) fit the Gompertz-gamma model to US Medicare data for both sexes by cohort during the period 1968-78. Their \hat{p} estimates for cohorts born between 1884 and 1902 lie in the range 0.0734-0.1029. The $\hat{\theta}$ estimates are published for the 1892 birth cohort (ages 75-85):

$$\begin{cases} \hat{\theta} = 4.7393 \text{ for males} \\ \hat{\theta} = 3.4722 \text{ for females} \end{cases}$$

Congdon (1994) uses the Gompertz-inverse Gaussian model and finds a slight improvement in fit in relation to the Gompertz-gamma model: this can be compared with the opposite finding of the study of the Manton et al. (1986) and this report. Further, Congdon experiments with the more general model of Aalen (1988) and finds that the inverse Gaussian frailty distribution is marginally preferable in this case.

6. DISCUSSION

Various models have been proposed for measuring heterogeneity in mortality data, based on continuous and discrete age formulations. We have considered

Vaupel et al. (1979)'s proportional frailty model because it is more straightforward to implement and to interpret the results, than the discrete alternatives. In this context, we have investigated, using two extensive insurance-based data sets, two frailty distributions, namely the gamma and the inverse Gaussian with respect to the Gompertz/Makeham family of individual hazard functions. The effects on the outcomes of using different graduation techniques have also been studied.

Regarding the assumptions for the frailty distribution, the analysis shows that the model incorporating the inverse Gaussian frailty distribution (GI) is less responsive to variations in the data and also is less likely to lead to satisfactory predictions for human mortality at the older ages (being an exponentially increasing function with age) than alternatives. In contrast, the models assuming a gamma frailty distribution (MP or P) are more favourable because of a better fit and because they have a logistic form with respect to age. Thus, when applied together with the Gompertz/Makeham individual hazard functions, the gamma frailty distribution seems to be more viable than the inverse Gaussian. This also predicts a non-homogeneous population for older ages, as discussed by Manton and Stallard (1984).

The generalized linear modelling method has a stronger theoretical justification than the classical non-linear least squares method and we find that it results in more stable parameter estimates. However, there are potential drawbacks when applying the GLM fitting technique to models with non-linear parameters. Firstly, one needs to know in advance the most likely domain for the non-linear parameters and secondly, the fitting process can fail to produce a global minimum of the deviance profile for the respective domain. We note that, in the particular case of the P model, the first disadvantage is usually not too restrictive, as we can often estimate permissible intervals for the a parameter. Further, this disadvantage is even less serious, when compared with the traditional method of non-linear regression, where we need good starting values for all of the parameters involved. Although the second drawback seems to be more serious in terms of fitting some of the data sets investigated, it should be noted that, for all such cases, we encountered difficulties with the NLM method too. Also, we note that the P models fitted by the weighted NLM technique proved to be strongly influenced by outliers in the data sets, often yielding, for example, estimates of \hat{a} that are too low compared to the observed forces of mortality (where we recall that $a = \lim_{x \rightarrow \infty} \mu_x$). We can also confirm the experience of other authors, that, when applied to the older ages only, the MP model is inferior to the P model, as the \hat{a} estimates do not appear to be significantly different from zero.

Regarding the level of heterogeneity in the investigated data sets, the results are less conclusive. The coefficient of variation of the gamma distribution ($1/\sqrt{\delta}$) for the final models is on average less than 0.45, implying 'fairly' homogeneous populations (corresponding to a $\hat{\delta}$ value of approximately 5). Making use of the $\hat{\delta} = \hat{a}/\hat{p}$ estimator of the P model (see Butt and Haberman 2002) we can attempt to draw up the potential range of the gamma shape parameter, based on this model. Parameter values corresponding to \hat{a} have been found in the literature in the range of 0.15 to 9.5. However, considering the interpretation of a , we probably can restrict this range to the interval from 0.35 to 1.3. For the

relative rate of increase of the force of mortality \hat{p} , the estimates are fairly stable, lying between 0.09 to 0.12. Hence, we have the following potential range:

$$\hat{\delta} = \frac{\hat{a}}{\hat{p}} = \frac{0.35}{0.12} \rightarrow \frac{1.3}{0.09} \simeq 2.916 \rightarrow 14.444 .$$

It is possible that the moderately high values of $\hat{\delta}$, observed above, are attributable to the investigated populations being shaped by selection (temporary initial selection for assured lives and adverse selection by annuitants: see Benjamin and Pollard 1993). For such populations, the basic frailty models, with constant frailty at the individual level during a life span, might not be sufficiently sensitive to pick up differences between the individual members. Then, a natural extension would be to study more advanced models where frailty varies with other factors, including age and physiological and socio-economic variables.

An implication of the presence of heterogeneity is that the life insurance company should seek more information about its policyholders when they purchase life insurance contracts, annuities or pensions, so that the relevant covariates can be identified and measured. This would then replicate the practice in non-life insurance (e.g. motor insurance) where a wide range of covariates are measured and used in the risk assessment. For life insurance contracts, insurers currently seek information on health status including smoking. For annuities and pensions, there has traditionally been little risk assessment in terms of identifying those who have a propensity to longevity (i.e. high risk for the insurers) but in some countries (eg the UK) there have been developments regarding impaired life annuities (see Ainslie 2000). The possibility of further and more detailed risk classification touches on political and ethical, as well as insurance, issues.

Overall, we conclude that the fitting of the Gompertz-gamma model (P) has been broadly successful for the two data sets discussed i.e. female annuitants and male assured lives. For these two data sets, the detailed results indicate that it would be worth investigating modelling the time trends in these parameter values with a view to developing methods for forecasting using time series methods (as in McNown and Rogers 1989) or incorporating the time parameter directly in the frailty model as advocated by Manton et al. (1981) and Vaupel (1999). This will be the subject of future research.

Finally, we remind readers of the identifiability discussion of section 5.1 and the need for caution in interpreting the meaning of the parameters in any frailty-based model. Thus, as noted earlier, the success in fitting the frailty-based model does not necessarily mean that the concepts underpinning the model are valid. Hence, it is possible that other concepts leading to the same model are correct. To proceed further with the diagnosis, further specific information is needed on the individuals whose experience is being analyzed.

ACKNOWLEDGMENTS

A special thanks should go to our colleague Arthur Renshaw for the invaluable comments and insights into the GLM technique during the preparation

of this work. Also the authors are grateful to the referees for their insightful comments which has led to an improvement to the presentation of the paper and to the CMI Bureau for providing the data. This research work was funded by a research grant from the Institute and Faculty of Actuaries, and by a research grant from the CMI Bureau.

APPENDIX

We consider the Gompertz model for μ_x i.e.

$$\mu_x = \beta \cdot e^{p \cdot x}$$

(Similar considerations would apply to the Gompertz-Makeham model $\mu_x = \alpha + \beta \cdot e^{p \cdot x}$).

If we choose a gamma distribution for $f_0(z)$, so that

$$f_0(z) = \frac{\theta^\delta \cdot z^{\delta-1} \cdot e^{-\theta z}}{\Gamma(\delta)}$$

then it is straightforward to show, using (2), that

$$\bar{\mu}(x) = \frac{\delta \cdot p \cdot \beta \cdot e^{p \cdot x}}{(\theta \cdot p - \beta) + \beta \cdot e^{p \cdot x}}$$

where we choose $\delta = \theta$ so that $\bar{z}_0 = 1$. An appropriate reparameterization leads to (5).

For the inverse Gaussian distribution, represented as

$$f_0(z) = \left[\frac{\psi}{\pi z^3} \right]^{\frac{1}{2}} \exp(4\psi\gamma)^{\frac{1}{2}} \exp(-\gamma z - \psi z^{-1})$$

then it is similarly straightforward to show that

$$\bar{\mu}(x) = e^{p \cdot x} \left(\frac{\psi \cdot \beta^2 \cdot p}{(\gamma \cdot p - \beta) + \beta \cdot e^{p \cdot x}} \right)^{\frac{1}{2}}$$

where we choose $\psi = \gamma$ so that $\bar{z}_0 = 1$. An appropriate reparameterization leads to (7). Here, as the anonymous referee has pointed out, we need to assume $\gamma \cdot p \neq \beta$, otherwise the model reduces to the basic Gompertz model. This condition is satisfied for real finite estimates of \hat{b} and \hat{p} obtained for model (7).

REFERENCES

- AALLEN, O.O. (1988) Heterogeneity in survival analysis, *Statistics in Medicine* **7**, 1121-1137.
 AINSLIE, R. (2000) Annuity and insurance products for impaired lives. Presented to the *Staple Inn Actuarial Society, Institute of Actuaries*, London.

- BENJAMIN, B. and POLLARD, J.H. (1993) *The analysis of mortality and other actuarial statistics*, The Institute of Actuaries and the Faculty of Actuaries, London.
- BLUMEN, I., KOGAN, M. and MCCARTHY, P.J. (1955) The industrial mobility of labour as a probability process, *Cornell Studies of Industrial and Labour Relations*, Vol. vi, Cornell University Press, Ithaca, New York.
- BUTT, Z. and HABERMAN, S. (2002) Application of frailty-based mortality models to insurance data, Actuarial Research Paper No. 142, Department of Actuarial Science & Statistics, City University, London.
- CONGDON, P. (1994) Analyzing mortality in London: Life Tables with frailty, *Journal of the Royal Statistical Society, Series D*, **43**, 277-308.
- FORFAR, D.O. and SMITH, D.M. (1988) The changing shape of English Life Tables, *Transactions of the Faculty of Actuaries*, **40**, 98-134.
- GINI, C. (1924) Premières recherches sur la fécondabilité de la femme, *Proceedings of the International Mathematical Congress*, **2**, 889-892.
- HABERMAN, S. and RENSHAW, A.E. (1996) Generalised linear models and actuarial science, *Journal of the Royal Statistical Society, Series D*, **45**, 407-436.
- HOEM, J.M. (1990) Identifiability in hazard model with unobserved heterogeneity: The compatibility of two apparently contradicting results, *Theoretical Population Biology* **37**, 124-128.
- HORIUCHI, S. and COALE, A.J. (1990) Age patterns of mortality for older women: An analysis using the age-specific rate of mortality change with age, *Mathematical Population Studies* **2**, 245-267.
- HOUGAARD, P. (1984) Life Table methods for heterogeneous populations: Distributions describing heterogeneity, *Biometrika* **71**, 75-83.
- KEYFITZ, N. (1985) *Applied mathematical demography*, Second Edition, Springer, New York.
- LANCASTER, T. (1990) *The econometric analysis of transition data*, Cambridge University Press, Cambridge.
- LE BRAS, H. (1976) Lois de mortalité et age limité, *Population* **31**, 655-692.
- MANTON, K.G. and STALLARD, E. (1984) Recent trends in mortality analysis, *Academic press, Inc.*, 236-299.
- MANTON, K.G., STALLARD, E. and VAUPEL, J.W. (1981) Methods for comparing the mortality experience of heterogeneous populations, *Demography* **18**, 389-410.
- MANTON, K.G., STALLARD, E. and VAUPEL, J.W. (1986) Alternative models for the heterogeneity of mortality risks among the aged, *Journal of the American Statistical Association* **81**, 635-644.
- MCCULLAGH, P. and NELDER, J.A. (1989) *Generalized linear models*, Second Edition, Chapman and Hall, London.
- MCKNOW, R. and ROGERS, A. (1989) Time-series analysis forecasts of a parameterised mortality schedule, in P. Congdon and P. Batey, eds, "Advances in regional demography: Information, forecasts, models", Bellhaven Press, London, pp. 107-123.
- RENSHAW, A.E. (1991) Actuarial graduation practice and generalised linear and non-linear models, *Journal of the Institute of Actuaries* **118**, 295-312.
- RENSHAW, A.E. (1995) Graduation and generalized linear models: An overview, Actuarial Research Paper No. 73, Department of Actuarial Science & Statistics, City University, London.
- THATCHER, A.R. (1999) The long-term pattern of adult mortality and the highest attained age, *Journal of the Royal Statistical Society, Series A*, **162**, 5-30.
- VAUPEL, J.W. (1999) Contribution to the discussion of 'The long-term pattern of adult mortality and highest attained age', by A.R. Thatcher, *Journal of the Royal Statistical Society, Series A*, **162**, 31-32.
- VAUPEL, J.W., MANTON, K.G. and STALLARD, E. (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography*, **16**, 439-454.
- VAUPEL, J.W. and YASHIN, A.I. (1985) The deviant dynamics of death in heterogeneous populations, *Sociological Methodology*, 179-211. Edited by N.B Tuma, Jossey-Bas, San Francisco.
- WANG, S.S. and BROWN, R.L. (1998) A frailty model for projection of human mortality improvement, *Journal of Actuarial Practice* **6**, 221-241.
- YASHIN, A.I., VAUPEL, J.W. and IACHINE, I.A. (1994) A duality in ageing: The equivalence of mortality models based on radically different concept, *Mechanisms of Ageing and Development* **74**, 1-14.