TRANSACTIONS OF SOCIETY OF ACTUARIES
1988 VOL. 40 PT 1

INCREASING AND INCREASING CONVEX BAYESIAN GRADUATION

JAMES D. BROFFITT

ABSTRACT*

It is well accepted that, for ages 30 and above, human mortality rates increase with age, yet no established graduation method makes direct use of this knowledge. Such a method is offered in this paper. Under the constant-force-of-mortality assumption, Bayes estimates (graduated values) are developed for the force of mortality. Smoothness of the graduated values is obtained by imposing one of two possible restrictions on the model — either that the force of mortality is increasing or that it is increasing and convex. This is accomplished by defining a prior distribution that assigns all of its probability to the set of parameters that satisfies the restriction. The resulting graduated values will automatically satisfy whichever smoothness restriction was assumed in the model.

Readers interested in obtaining a computer program to do these graduations should send a self-addressed, stamped mailer and a 5.25-inch double-sided double-density diskette to the author at his Yearbook address. Diskettes will be returned with four files containing information, data used in this paper, FORTRAN 77 source code, and the corresponding compiled executable program, which will run on an IBM PC or compatible system.

I. INTRODUCTION

The construction of mortality tables traditionally has been viewed as a two-stage process: (1) initial mortality rates are calculated and (2) these rates are smoothed by using a graduation technique. Several graduation methods have been proposed, but they often are geared to smoothing the data with little or no consideration for a suitable statistical model of the process that generated the data. Although these techniques have served us well, in recent years it has been recognized that graduation may be formulated as a problem in the statistical estimation of unknown parameters.

*I give my thanks and praise to God for providing the tools to do this work.

Let $q_x, \ldots, q_{x+k-1}$ be the mortality rates at ages $x, \ldots, x+k-1$ for a particular population. We take the view that graduation is the process of *simultaneously* estimating the k unknown parameters $q_x, \ldots, q_{x+k-1}$. Because of the vast amount of information that is often available from previous studies, a Bayesian analysis is suggested as the most appropriate estimation method. Three items are needed for a Bayesian analysis: (1) a likelihood function, (2) a prior distribution for the unknown parameters in this likelihood, and (3) the observed data. These three elements are blended to obtain the conditional probability distribution of the parameters given the observed data, that is, the posterior distribution. The Bayes estimate is then taken as a measure of the "middle" of the posterior distribution. Readers wishing to review these concepts should consult a text on mathematical statistics such as Hogg and Craig [10].

Throughout this paper we make the simplifying assumption that the force of mortality, $\mu(y)$, is constant over unit age intervals. Thus for $j = 1, \ldots, k$

$$\mu(y) = \theta_j, \ x+j-1 \le y < x+j, \tag{1.1}$$

and therefore,

$$q_{x+j-1} = 1 - \exp(-\theta_j). \tag{1.2}$$

Notice that the subscript of q denotes age, while the subscript of θ indexes the age intervals.

We find it convenient to take $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ as the basic parameter to be estimated, and in this paper Bayes estimates will be developed for $\boldsymbol{\theta}$. Then the mortality rate q_{x+j-1} is estimated by substituting the estimate of θ_j into (1.2).

One goal of graduation is to obtain a "smooth" sequence of estimates. In a statistical approach this can be accomplished by including smoothness assumptions as a part of the model. These assumptions should characterize smoothness adequately, but also should be simple enough to make computation of the estimates feasible.

We shall obtain smoothness by restricting $\boldsymbol{\theta}$ to lie within certain subsets of the full parameter space, $\Omega = \{\boldsymbol{\theta};\ 0 \le \theta_i < \infty,\ i = 1, \ldots, k\}$. The two different subsets that will be considered correspond to characteristics displayed by the force of mortality. Let $R_I = \{\boldsymbol{\theta};\ 0 < \theta_1 < \ldots < \theta_k\}$. Assumption (or restriction) (I) designates $\boldsymbol{\theta} \in R_I$. Such $\boldsymbol{\theta}$'s will be called "increasing." If $\boldsymbol{\theta}$ is increasing, it follows that $q_x < \ldots < q_{x+k-1}$. It is well accepted that human mortality rates increase with age, after some pivotal age in the late twenties. Thus we should be safe in applying restriction

(I) for $x \geq 30$. Graduations over this age range have sometimes been criticized for not providing increasing mortality rates. Restriction (I) will automatically produce graduations that are increasing.

Additional smoothness may be obtained by further restricting $\boldsymbol{\theta}$ to lie in the set

$$R_{IC} = \{\boldsymbol{\theta};\ 0 < \theta_1 < \infty,\ 0 < \theta_2 - \theta_1 < \ldots < \theta_k - \theta_{k-1}\},$$

which is a subset of R_I. If $\boldsymbol{\theta} \in R_{IC}$, then, in addition to being increasing, the elements of $\boldsymbol{\theta}$ have increasing increments or first differences. Such $\boldsymbol{\theta}$'s will be called "increasing convex." The restriction $\boldsymbol{\theta} \in R_{IC}$ will be referred to as assumption (IC). As with assumption (I), assumption (IC) should not be imposed over age ranges that dip below 30.

If the upper limit of life, ω, is finite, then by necessity, $\mu(y) \rightarrow +\infty$ as $y \rightarrow \omega$. Restriction (IC) simply forces $\mu(y)$ to approach $+\infty$ in a smooth convex fashion. A simple illustration is given by de Moivre's law, where $\mu(y) = 1/(\omega - y)$.

Although the mortality rates must be increasing when $\boldsymbol{\theta} \in (IC)$, they may or may not be convex. For example, let $k = 3$ and notice that $\boldsymbol{\theta}_1 = (0.10, 0.20, 0.31)$ and $\boldsymbol{\theta}_2 = (0.10, 0.20, 0.32)$ belong to R_{IC}. Corresponding to $\boldsymbol{\theta}_1$, $\Delta^2 q_x < 0$, but for $\boldsymbol{\theta}_2$, $\Delta^2 q_x > 0$, where q_x, q_{x+1}, and q_{x+2} are computed according to (1.2).

Let $\boldsymbol{\theta} \in R$ denote a general restriction. If we believe $\boldsymbol{\theta} \in R$, then the prior distribution of $\boldsymbol{\theta}$ should assign probability one to the set R. In section III we shall define prior distributions for which $P[\boldsymbol{\theta} \in R] = 1$. It will then follow that $P[\boldsymbol{\theta} \in R|\text{Data}] = 1$, that is, the posterior distribution also has support R. If R is a convex set (both R_I and R_{IC} are convex sets), any reasonable measure of central tendency of the posterior distribution also will belong to R. We may then be assured that our graduated value of $\boldsymbol{\theta}$ will be in R and thus satisfy the smoothness requirement.

The estimation of monotone parameters is usually called isotonic estimation. Much research has been done on this problem from a non-Bayesian viewpoint. The interested reader is referred to the book by Barlow et al. [1]. Hildreth [9] and Dent and Robertson [7] have considered the estimation of increasing concave parameters.

Previous research in Bayesian estimation of increasing parameters may be found in Smith [12] and Broffitt [3–5]. Readers also may wish to refer to the Bayesian graduation technique developed by Kimeldorf and Jones [11].

Their analysis was based on multivariate normal distributions, and they obtained smoothness by specifying large positive correlation coefficients among neighboring mortality rates in the prior distribution. Some refinements of this procedure were suggested by Hickman and Miller [8].

II. THE LIKELIHOOD FUNCTION

We are interested in mortality rates between ages x and $x+k$. Suppose N lives come under observation at some point within this age interval. Let $x+s_i$ be the age at which observation of life i begins, and $x+t_i$ be the age at which observation ceases. This cessation of obscrvtion may be caused by death, voluntary withdrawal, the attainment of age $x+k$, or termination of the observation period. Under a model that assumes independent random times to death and withdrawal, the likelihood function, L, is given by

$$L \propto \prod_{i \in \mathcal{D}} \mu(x+t_i) \cdot \exp\left[-\sum_{i=1}^{N} \int_{x+s_i}^{x+t_i} \mu(y)dy\right], \tag{2.1}$$

where $\mathcal{D}$ denotes the subset of subscripts corresponding to those lives that died and $A \propto B$ denotes that A is proportional to B, that is, $A = cB$. (For further elaboration, see Broffitt [2, 3], Steelman [13], and Chan and Panjer [6].)

Assumption (1.1) may be utilized to simplify (2.1). Letting d_j equal the number of deaths observed between ages $x+j-1$ and $x+j$,

$$\prod_{i \in \mathcal{D}} \mu(x + t_i) = \prod_{j=1}^{k} \theta_j^{d_j}.$$

Also, let e_{ij} equal the amount of time, measured in years, that life i was under observation between ages $x+j-1$ and $x+j$. Then

$$\begin{aligned}\sum_{i=1}^{N}\left[\int_{x+s_i}^{x+t_i} \mu(y)dy\right] &= \sum_{i=1}^{N}\sum_{j=1}^{k} e_{ij}\,\theta_j \\ &= \sum_{j=1}^{k} e_j\theta_j,\end{aligned}$$

where $e_j = \sum_{i=1}^{N} e_{ij}$ is the total number of years the N lives were under observation between ages $x+j-1$ and $x+j$. Thus (2.1) simplifies to

$$L(\boldsymbol{\theta}) \propto \prod_{j=1}^{k} \left[\theta_j^{d_j} e^{-e_j \theta_j} \right]. \tag{2.2}$$

The unrestricted maximum likelihood estimator of θ_j is $\theta_j^M = d_j/e_j$, which corresponds to the well-known result in reliability theory that when lifetime is exponentially distributed, the hazard rate is estimated by total failures divided by total time on test. In our problem e_j will be referred to as "exposure."

Finally, we note that the Bayes estimators to be developed are not particular to the mortality rate problem, but apply whenever the likelihood has the same form as (2.2). For example, suppose $\theta_1, \ldots, \theta_k$ are the parameters for k independent Poisson distributions. If a sample of size e_j is taken from the jth distribution and the sample total is d_j, then the likelihood is given by (2.2).

III. PRIOR DISTRIBUTIONS AND ESTIMATORS

The following notation will be useful: $g(x|a,b)$ will denote the gamma probability density function (pdf), $b^a x^{a-1} \exp(-bx)/\Gamma(a)$, for $x > 0$, and $g(\mathbf{x}|\mathbf{a},\mathbf{b})$ will denote $\prod_{i=1}^{k} g(x_i|a_i,b_i)$, where $\mathbf{x} = (x_1, \ldots, x_k)$, $\mathbf{a} = (a_1, \ldots, a_k)$, and $\mathbf{b} = (b_1, \ldots, b_k)$. We shall use prior(·) and post(·) to denote the prior and posterior pdf's. Also, let $\mathbf{u}^{(i)}$ be a k-dimensional vector with a 1 in the ith position and zeros elsewhere.

Conjugate Prior

For a general restriction, $\boldsymbol{\theta} \in R$, consider the prior pdf given by

$$\begin{aligned} \text{prior}(\boldsymbol{\theta}) &= g(\boldsymbol{\theta}|\boldsymbol{\delta},\boldsymbol{\epsilon})/\rho(R), \quad \boldsymbol{\theta} \in R \\ &= 0, \qquad \text{elsewhere,} \end{aligned} \tag{3.1}$$

where the scaling constant is $\rho(R) = \int_R g(\boldsymbol{\theta}|\boldsymbol{\delta},\boldsymbol{\epsilon})d\boldsymbol{\theta}$. Combining (2.2) and (3.1), we have

$$\text{post}(\boldsymbol{\theta}) \propto \text{L}(\boldsymbol{\theta}) \cdot \text{prior}(\boldsymbol{\theta}),$$

and thus

$$\text{post}(\boldsymbol{\theta}) = g(\boldsymbol{\theta}|\boldsymbol{\alpha},\boldsymbol{\lambda})/\text{p}(R), \quad \boldsymbol{\theta} \in R \qquad (3.2)$$
$$= 0, \qquad \text{elsewhere,}$$

where $\alpha_j = \delta_j + d_j$, $\lambda_j = \epsilon_j + e_j$, and $\text{p}(R) = \int_R g(\boldsymbol{\theta}|\boldsymbol{\alpha},\boldsymbol{\lambda})d\boldsymbol{\theta}$. By comparing (3.1) and (3.2), it is clear that (3.1) is the conjugate prior pdf.

If the Bayes estimator is the posterior mean, denoted by $\boldsymbol{\theta}^B(R) = (\theta_1^B(R), \ldots, \theta_k^B(R))$, then

$$\theta_i^B(R) = \int_R \theta_i\, g(\boldsymbol{\theta}|\boldsymbol{\alpha},\boldsymbol{\lambda})d\boldsymbol{\theta}/\text{p}(R). \qquad (3.3)$$

Since $\theta g(\theta|\alpha,\lambda) = (\alpha/\lambda)g(\theta|\alpha+1,\lambda)$, (3.3) reduces to

$$\theta_i^B(R) = \theta_i^B \cdot [\text{p}^{(i)}(R)/\text{p}(R)], \qquad (3.4)$$

where $\theta_i^B = \theta_i^B(\Omega) = \alpha_i/\lambda_i$ is the unrestricted Bayes estimator of θ_i, $\boldsymbol{\alpha}^{(i)} = \boldsymbol{\alpha} + \mathbf{u}^{(i)}$ and $\text{p}^{(i)}(R) = \int_R g(\boldsymbol{\theta}|\boldsymbol{\alpha}^{(i)},\boldsymbol{\lambda})d\boldsymbol{\theta}$.

The fundamental result given in (3.4) expresses the Bayes estimator $\theta_i^B(R)$ in a seemingly simple form. It is obtained by multiplying the unrestricted Bayes estimator by a ratio of "similar" probabilities. In applications, $\text{p}(R)$ and, consequently, $\text{p}^{(i)}(R)$ can be quite difficult to compute. For assumption (I), FORTRAN programs for calculating $\text{p}(R_I)$ were provided by Broffitt [3, 4]; however, the CPU time required can become prohibitively large as k and $\alpha_1, \ldots, \alpha_k$ increase. In the earlier paper, selection of the prior parameters $\boldsymbol{\delta}$ and $\boldsymbol{\epsilon}$ was by trial and error, which also required time-consuming computations. For these reasons we shall abandon the conjugate prior in this paper and opt for what previously has been called the additive prior. Although it is possible to compute the mean of the resulting posterior, this also poses computational difficulties; hence we shall shift from the posterior mean to the posterior mode. With these modifications, the calculations become extremely fast and the graduations appear to be quite good.

Additive Prior: Restriction (I)

Under restriction (I) we find it convenient to reparameterize. Let

$$\begin{aligned}\theta_1 &= \phi_1 \\ \theta_2 &= \phi_1 + \phi_2 \\ &\ \ \vdots \\ \theta_k &= \phi_1 + \phi_2 + \ldots + \phi_k.\end{aligned} \tag{3.5}$$

The parameters $\phi_2, \ldots, \phi_k$ represent increments in the θ's. Clearly the θ's are increasing if and only if their increments or first differences are positive. The transformation (3.5) is one-to-one between the sets $\{\boldsymbol{\theta}; 0 < \theta_1 < \ldots < \theta_k < \infty\}$ and $\{\boldsymbol{\phi}; 0 < \phi_i < \infty, i = 1, \ldots, k\}$, and the Jacobian of this transformation is 1.

The importance of considering the Jacobian may be seen in the following result. Suppose $f_{\mathbf{X}}(\mathbf{x})$ is the pdf of the random vector $\mathbf{X}$, and we wish to find the pdf of the random vector $\mathbf{Y}$, which is given by $\mathbf{Y} = \mathbf{g}(\mathbf{X})$. If the transformation from $\mathbf{X}$ to $\mathbf{Y}$ is one-to-one, so that $\mathbf{X} = \mathbf{g}^{-1}(\mathbf{Y})$, and J is the Jacobian of the transformation, then the pdf of $\mathbf{Y}$ is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y}))|J|.$$

(See Hogg and Craig [10, section 4.5].) If the Jacobian is a constant function of $\mathbf{y}$, then the pdf of $\mathbf{Y}$ is proportional to the pdf of $\mathbf{X}$.

In our problem $J = 1$, so that the pdf of $\boldsymbol{\theta}$, which may represent either a prior or posterior distribution, may be obtained from the pdf of $\boldsymbol{\phi}$ as

$$f_{\boldsymbol{\theta}}(\theta_1, \ldots, \theta_k) = f_{\boldsymbol{\phi}}(\theta_1, \theta_2 - \theta_1, \ldots, \theta_k - \theta_{k-1}).$$

Also, the pdf of $\boldsymbol{\phi}$ may be obtained from that of $\boldsymbol{\theta}$ as

$$f_{\boldsymbol{\phi}}(\phi_1, \ldots, \phi_k) = f_{\boldsymbol{\theta}}(\phi_1, \phi_1 + \phi_2, \ldots, \phi_1 + \ldots + \phi_k).$$

Thus we can work with either set of parameters and easily switch from one to the other by simply substituting according to (3.5).

The procedure is first to find the Bayes estimate of $\boldsymbol{\phi}$ and then use (3.5) to calculate the corresponding estimate of $\boldsymbol{\theta}$.

Let the prior pdf of ϕ be given by

$$\text{prior}(\phi) = g(\phi|\mathbf{a},\mathbf{r}).$$

From (2.2) and (3.5),

$$\begin{aligned}\text{post}(\phi) &\propto \prod_{j=1}^{k} \theta_j^{d_j} e^{-e_j\theta_j} \cdot g(\phi|\mathbf{a},\mathbf{r}) \\ &\propto \prod_{j=1}^{k} \theta_j^{d_j} \phi_j^{a_j-1} e^{-b_j\phi_j}, \ \phi_j > 0, \ j = 1, \ldots, k,\end{aligned} \tag{3.6}$$

where $\theta_j = \phi_1 + \ldots + \phi_j$ and $b_j = r_j + e_j + \ldots + e_k$.

Under assumption (I), let $\phi^B(I) = (\phi_1^B(I), \ldots, \phi_k^B(I))$ and $\theta^B(I) = (\theta_1^B(I), \ldots, \theta_k^B(I))$ denote the Bayes estimates of ϕ and θ. We take $\phi^B(I)$ to be the mode of post(ϕ), and according to (3.5), we define $\theta_j^B(I) = \sum_{i=1}^{j} \phi_i^B(I)$. While $\phi^B(I)$ is the posterior mode of ϕ, $\theta^B(I)$ is not the posterior mode of θ. Rather, it is a transformation of the posterior mode of ϕ according to the reparameterization (3.5).

On examination of (3.6), it is evident that if any of $d_1+a_1, a_2, \ldots, a_k$ is less than 1, then the posterior mode is on the boundary of the support of post(ϕ). For example, if $a_2 < 1$, post(ϕ) can be made arbitrarily large by taking ϕ_2 close enough to zero. In order to avoid this undesirable situation, we shall require $a_i > 1, i = 1, \ldots, k$.

The estimate $\phi^B(I)$ is obtained by solving

$$\frac{\partial}{\partial\phi_i} \ln \text{post}(\phi) = 0, \ i = 1, \ldots, k.$$

This system of equations reduces to

$$\sum_{j=i}^{k} \frac{d_j}{\theta_j} + \frac{a_i - 1}{\phi_i} - b_i = 0, \ i = 1, \ldots, k, \tag{3.7}$$

where $\theta_j = \phi_1 + \ldots + \phi_j$. The solution is quite easy to find by using a numerical iteration procedure. Examples are provided in Section V, and the iteration technique is described in the Appendix.

Additive Prior: Restriction (IC)

Under restriction (*IC*) the increments of the θ's must be increasing as well as positive. For this case the reparameterization is:

$$
\begin{aligned}
\theta_1 &= \psi_1 \\
\theta_2 &= \psi_1 + \psi_2 \\
\theta_3 &= \psi_1 + 2\psi_2 + \psi_3 \\
\theta_4 &= \psi_1 + 3\psi_2 + 2\psi_3 + \psi_4 \qquad (3.8)\\
&\ \ \vdots \\
\theta_k &= \psi_1 + (k-1)\psi_2 + (k-2)\psi_3 + \ldots + \psi_k.
\end{aligned}
$$

The increments of the θ's are

$$\nabla\theta_j = \psi_2 + \ldots + \psi_j,\ j = 2, \ldots, k,$$

which are positive and increasing as long as $\psi_j > 0$, $j = 2, \ldots, k$. The transformation given by (3.8) is one-to-one between the sets

$$\{\boldsymbol{\theta};\ 0 < \theta_1 < \infty, 0 < \theta_2 - \theta_1 < \ldots < \theta_k - \theta_{k-1} < \infty\}$$

and $\{\boldsymbol{\psi};\ 0 < \psi_i < \infty,\ i = 1, \ldots, k\}$, and, as before, the Jacobian is 1. We may now proceed in a manner analogous to that used for assumption (*I*).

Let the prior pdf of $\boldsymbol{\psi}$ be given by prior($\boldsymbol{\psi}$) = $g(\boldsymbol{\psi}|\mathbf{a},\mathbf{r})$. Then

$$
\begin{aligned}
\text{post}(\boldsymbol{\psi}) &\propto \prod_{j=1}^{k} \theta_j^{d_j} e^{-e_j\theta_j} \cdot g(\boldsymbol{\psi}|\mathbf{a},\mathbf{r}) \\
&\propto \prod_{j=1}^{k} \theta_j^{d_j} \psi_j^{a_j-1} e^{-c_j\psi_j},\ \psi_j > 0,\ j = 1, \ldots, k, \qquad (3.9)
\end{aligned}
$$

where

$$\theta_j = \psi_1, \qquad j = 1$$

$$= \psi_1 + \sum_{i=2}^{j} (j-i+1)\psi_i, \; j = 2, \ldots, k,$$

and

$$c_j = r_1 + e_1 + \ldots + e_k, \; j = 1$$

$$= r_j + \sum_{i=j}^{k} (i-j+1)e_i, \quad j = 2, \ldots, k.$$

The posterior mode, $\boldsymbol{\psi}^B(IC)$, is obtained by simultaneously solving

$$\frac{\partial}{\partial\psi_i} \ln \text{post}(\boldsymbol{\psi}) = 0, \; i = 1, \ldots, k.$$

This system reduces to

$$\sum_{j=1}^{k} (d_j/\theta_j) + (a_1 - 1)/\psi_1 - c_1 = 0, \qquad i = 1$$

$$\sum_{j=i}^{k} (d_j/\theta_j)(j-i+1) + (a_i - 1)/\psi_i - c_i = 0, \qquad i = 2, \ldots, k. \tag{3.10}$$

For the reason cited in the discussion of assumption (I), we again shall use only values of $\mathbf{a}$ for which $a_i > 1$, $i = 1, \ldots, k$. After obtaining $\boldsymbol{\psi}^B(IC)$, (3.8) is used to compute the estimate of $\boldsymbol{\theta}$, $\boldsymbol{\theta}^B(IC)$, as

$$\theta_j^B(IC) = \psi_1^B(IC), \qquad j = 1$$

$$= \psi_1^B(IC) + \sum_{i=2}^{j} (j-i+1)\psi_i^B(IC), \qquad j = 2, \ldots, k. \tag{3.11}$$

IV. SELECTING THE PRIOR PARAMETERS

Before computing the Bayes estimates, it is necessary to specify values for the prior parameters $\mathbf{a}$ and $\mathbf{r}$. We begin by considering characteristics of location and spread of the prior distribution of $\boldsymbol{\theta}$. These will be used to determine $\mathbf{a}$ and $\mathbf{r}$.

Let $\boldsymbol{\theta}^P = (\theta_1^P, \ldots, \theta_k^P)$ be our prior guess for the value of $\boldsymbol{\theta}$. If the restriction is $\boldsymbol{\theta} \in R$, then the value selected for $\boldsymbol{\theta}^P$ must be a point in R. In practice $\boldsymbol{\theta}^P$ generally will be based on a previous graduation or an established mortality table. Slight adjustments of such tabulated values might be required to ensure that $\boldsymbol{\theta}^P \in R$.

Let $v_i^P = \text{Var}(\theta_i)$ and $\mathbf{v}^P = (v_1^P, \ldots, v_k^P)$. While specification of $\boldsymbol{\theta}^P$ defines the level of mortality expressed by prior knowledge, specifying $\mathbf{v}^P$ defines the precision of this prior knowledge. Decreasing the values in $\mathbf{v}^P$ implies a greater confidence in prior information, so the graduation, that is, $\boldsymbol{\theta}^B(\cdot)$, should be drawn closer to $\boldsymbol{\theta}^P$.

The elements of $\mathbf{v}^P$ also must satisfy certain relationships. For restriction (I), we must choose $\mathbf{v}^P$ so that $v_1^P < \ldots < v_k^P$. This is clear from (3.5) and the fact that the elements of $\boldsymbol{\phi}$ are stochastically independent in the prior distribution. This does not seem to be a severe restriction because it is natural for the variablity of θ_i to increase as its expectation increases. Also, we have less experience for the very old ages, and thus our prior information is probably less certain. Accordingly, we should insist that the prior variances be larger at these older ages.

Under assumption (IC), the requirement for $\mathbf{v}^P$ is more stringent. This is caused by the coefficients of the ψ's in the reparameterization (3.8). A simple example illustrates this point. Let $k = 3$, and suppose we try $\mathbf{v}^P = (0.010, 0.011, 0.013)$. Then $\text{Var}(\psi_1) = 0.010$, $\text{Var}(\psi_2) = 0.001$, and $\text{Var}(\psi_3) = -0.001$ which, of course, is inadmissible. The point is that we must be cautious in selecting $\mathbf{v}^P$. The values chosen must be such that $\psi_1, \ldots, \psi_k$ are legitimate random variables.

Let $v_i^M = \text{Var}(\theta_i^M)$, where $\theta_i^M = d_i/e_i$ is the unrestricted maximum likelihood estimator of θ_i. One way to achieve balance between the precision of the prior information and the precision of the data is to set $v_i^P = m \cdot v_i^M$. The value of m must be specified by the graduator. Increasing m places more emphasis on the data, and decreasing m places more emphasis on prior knowledge.

For v_i^M we shall use the approximation (see Broffitt [4, p. 30])

$$v_i^M \doteq [\exp(\theta_i^P) - 1]/e_i. \tag{4.1}$$

From (4.1) we see that setting $v_i^P = m v_i^M$ may violate the monotonicity constraint placed on $\mathbf{v}^P$ by either restriction (I) or (IC). That is, in going from one age to the next, the increase in θ_i^P should be quite small but the variability in e_i may be substantial. Thus we should not expect that $v_1^M < \ldots < v_k^M$.

This difficulty may be resolved by merging the k conditions, $v_i^P = mv_i^M$, i = 1, . . . , k, into a single condition,

$$\sum_{i=1}^{k} v_i^P = m \sum_{i=1}^{k} v_i^M. \tag{4.2}$$

Together with specification of the value of θ_i^P, $i = 1, \ldots, k$, this provides $k+1$ conditions for determining **a** and **r**. In order to obtain a unique solution, we further simplify the problem by assuming $a_1 = \ldots = a_k$. This common value will be denoted by α.

First consider assumption (I) and let $\boldsymbol{\phi}^P = (\phi_1^P, \ldots, \phi_k^P)$ be defined by $\theta_j^P = \phi_1^P + \ldots + \phi_j^P$, $j = 1, \ldots, k$. Then $\boldsymbol{\phi}^P$ is the prior guess for $\boldsymbol{\phi}$ corresponding to the choice of $\boldsymbol{\theta}^P$ as the prior guess for $\boldsymbol{\theta}$. Because the Bayes estimate of $\boldsymbol{\phi}$ is the posterior mode, it seems logical to set $\boldsymbol{\phi}^P$ equal to the prior mode of $\boldsymbol{\phi}$, that is,

$$\phi_i^P = (\alpha - 1)/r_i, \; i = 1, \ldots, k.$$

Thus $\mathrm{Var}(\phi_i) = \alpha/r_i^2 = (\phi_i^P)^2\, \alpha/(\alpha - 1)^2$, and (4.2) becomes

$$\alpha/(\alpha-1)^2 \cdot \sum_{i=1}^{k} h_i(\phi_i^P)^2 = m \sum_{i=1}^{k} v_i^M, \tag{4.3}$$

where $h_i = k - i + 1$. Now solve (4.3) to obtain

$$\alpha = 1 + u + \sqrt{u(2 + u)}, \tag{4.4}$$

where

$$u = \sum_{i=1}^{k} h_i(\phi_i^P)^2 / \left[2m \sum_{i=1}^{k} v_i^M \right]. \tag{4.5}$$

The sign of the square root in (4.4) is positive, because we require $\alpha > 1$. Finally r_i is calculated as

$$r_i = (\alpha - 1)/\phi_i^P, \; i = 1, \ldots, k. \tag{4.6}$$

Under assumption (*IC*) we may proceed in an analogous manner. In this case let $\boldsymbol{\psi}^P = (\psi_1^P, \ldots, \psi_k^P)$ be defined by

$$\begin{aligned} \theta_j^P &= \psi_1^P, & j &= 1 \\ &= \psi_1^P + \sum_{i=2}^{j} (j-i+1)\, \psi_i^P, & j &= 2, \ldots, k. \end{aligned}$$

Then set $\boldsymbol{\psi}^P$ equal to the prior mode of $\boldsymbol{\psi}$ so that $\psi_i^P = (\alpha - 1)/r_i$, $i = 1, \ldots, k$. Corresponding to (4.3) we obtain

$$\alpha/(\alpha - 1)^2 \cdot \sum_{i=1}^{k} h_i(\psi_i^P)^2 = m \sum_{i=1}^{k} v_i^M,$$

where, in this case,

$$\begin{aligned} h_i &= k, & i &= 1 \\ &= \sum_{j=1}^{k-i+1} j^2, & i &= 2, \ldots, k. \end{aligned}$$

Thus the solution for α is as given in (4.4) but with

$$u = \sum_{i=1}^{k} h_i(\psi_i^P)^2 \Big/ \left[2m \sum_{i=1}^{k} v_i^M \right], \tag{4.7}$$

and r_i calculated as

$$r_i = (\alpha - 1)/\psi_i^P,\ i = 1, \ldots, k. \tag{4.8}$$

After a suitable choice has been made for $\boldsymbol{\theta}^P$, the graduator need only select m and then calculate α (recall $\alpha = a_1 = \ldots = a_k$) and r_i, $i = 1, \ldots, k$. Under restriction (*I*) these steps are completed by using (4.5), (4.4), and (4.6). Under restriction (*IC*) the correct expressions are (4.7), (4.4), and (4.8). Care must be exercised to use the appropriate definition of h_i; it is different under restrictions (*I*) and (*IC*).

We should refrain from attaching too much significance to the numerical value of m. It is true that $m = 1$ places more emphasis on prior knowledge than does $m = 2$, but it is not clear how to measure the weight $m = 1$ places on prior knowledge relative to the data.

In an attempt to measure the weight of the data relative to prior knowledge, we define the statistic w as

$$w = \sum_{i=1}^{k} w_i/k,$$

where

$$w_i = |\theta_i^P - \theta_i^B(\cdot)|/[|\theta_i^P - \theta_i^B(\cdot)| + |\theta_i^B(\cdot) - \theta_i^M|].$$

(This measure was suggested by my colleague Robert V. Hogg.) Notice that w_i (and therefore w) is between 0 and 1. The closer $\theta_i^B(\cdot)$ is to θ_i^M, the closer w_i is to 1, and the closer $\theta_i^B(\cdot)$ is to θ_i^P, the closer w_i is to 0. If $\theta_i^B(\cdot)$ is halfway between θ_i^P, and θ_i^M, or if $\theta_i^P = \theta_i^M$, then $w_i = 1/2$. In summary, $w = 1/2$ indicates equal weighting between prior knowledge and data; $w < 1/2$ indicates more weight on prior knowledge; and $w > 1/2$ indicates more weight on the data.

Because of the nature of the restriction and the jagged pattern displayed by $\boldsymbol{\theta}^M$, it may be impossible for w to be greater than 1/2. For example, since under restriction (IC), $\boldsymbol{\theta}^B(IC)$ must be increasing convex, no matter how large m is made, it may be impossible to draw $\boldsymbol{\theta}^B(IC)$ "close" to $\boldsymbol{\theta}^M$. By assuming $\boldsymbol{\theta} \in R_{IC}$, we are imparting a good deal of prior knowledge into the estimation process, knowledge not about the level of mortality but rather about the pattern of mortality.

We conclude this section with a generalization of the technique described above for selecting **a** and **r**. Rather than having a common value, α, for each a_i, partition **a** into n groups and let the a's within the jth group have a common value, α_j. Let k_j be the number of a's in the jth group, so that $k_1 + \ldots + k_n = k$, and let $g_j = \sum_{i=1}^{j} k_i$, $j = 1, \ldots, n$, and $g_0 = 0$. Thus

$$\alpha_j = a_{g_{j-1}+1} = \ldots = a_{g_j}, j = 1, \ldots, n.$$

To specify the prior parameters **a** and **r**, we must specify $n+k$ values. This will be done by using the k conditions resulting from selecting $\boldsymbol{\theta}^P$, and the n conditions

$$\sum_{i=g_{j-1}+1}^{g_j} v_i^P = m_j \sum_{i=g_{j-1}+1}^{g_j} v_i^M, j = 1, \ldots, n. \qquad (4.9)$$

This method for selecting **a** and **r** provides the opportunity to assign different weights to the prior information over different age ranges.

Under assumption (*I*),

$$\sum_{i=g_{j-1}+1}^{g_j} v_i^P = \sum_{i=1}^{g_j} v_i^P - \sum_{i=1}^{g_{j-1}} v_i^P$$

$$= \sum_{i=1}^{g_j} h_{ij} \operatorname{Var}(\phi_i) - \sum_{i=1}^{g_{j-1}} h_{i,\,j-1} \operatorname{Var}(\phi_i)$$

$$= \sum_{i=1}^{g_{j-1}} (h_{ij} - h_{i,\,j-1}) \operatorname{Var}(\phi_i) + \alpha_j/(\alpha_j - 1)^2$$

$$\cdot \sum_{i=g_{j-1}+1}^{g_j} h_{ij} (\phi_i^P)^2, \tag{4.10}$$

where

$$h_{ij} = g_j - i + 1,\ i = 1, \ldots, g_j,\ j = 1, \ldots, n.$$

It will be convenient to use the following notation:

$$T_{1j} = \sum_{i=g_{j-1}+1}^{g_j} h_{ij} (\phi_i^P)^2, \qquad j = 1, \ldots, n$$

$$T_{2j} = \sum_{i=g_{j-1}+1}^{g_j} v_i^M, \qquad j = 1, \ldots, n$$

$$T_{3j} = 0, \qquad j = 1$$

$$= \sum_{i=1}^{g_{j-1}} (h_{ij} - h_{i,\,j-1}) \operatorname{Var}(\phi_i), \quad j = 2, \ldots, k.$$

Then, using (4.10), (4.9) becomes

$$T_{3j} + \alpha_j/(\alpha_j - 1)^2\, T_{1j} = m_j T_{2j},$$

and therefore,

$$\alpha_j = 1 + u_j + \sqrt{u_j\,(2 + u_j)},$$

where

$$u_j = T_{1j}/[2(m_jT_{2j} - T_{3j})].$$

Since $u_j = (\alpha_j - 1)^2/(2\alpha_j)$, u_j must be positive. This imposes a lower bound on m_j, that is, $m_j > T_{3j}/T_{2j}$. After obtaining α_j, r_i is computed as

$$r_i = (\alpha_j - 1)/\phi_i^P, i = g_{j-1} + 1, \ldots, g_j.$$

The values $m_1, \ldots, m_n$ must be selected sequentially. Once a value has been chosen for m_1, the quantities α_1, r_i, $i = 1, \ldots, g_1$, and $T_{3,2}$ are computed. Next a value for m_2 is chosen and α_2, r_i, $i = g_1 + 1, \ldots, g_2$, and $T_{3,3}$ are computed. This process continues until **a** and **r** have been determined.

Under restriction (*IC*) the same procedure is used but with the following definitions of h_{ij}, T_{1j}, T_{3j}, and r_i:

$$\begin{aligned} h_{ij} &= g_j, \qquad i = 1,\ j = 1, \ldots, n \\ &= \sum_{l=1}^{g_j - i + 1} l^2, \ i = 2, \ldots, g_j, \quad j = 1, \ldots, n \\ T_{1j} &= \sum_{i=g_{j-1}+1}^{g_j} h_{ij}(\psi_i^P)^2, \quad j = 1, \ldots, n \\ T_{3j} &= 0, \qquad j = 1 \\ &= \sum_{i=1}^{g_{j-1}} (h_{ij} - h_{i,j-1}) \operatorname{Var}(\psi_i), \quad j = 2, \ldots, k \\ r_i &= (\alpha_j - 1)/\psi_i^P, i = g_{j-1} + 1, \ldots, g_j, \ j = 1, \ldots, n. \end{aligned}$$

V. EXAMPLES

The examples are based on a set of data that represent male ultimate (duration $\geq$ 16) experience for premium-paying policies with face amounts between \$10,000 and \$24,900. (I am indebted to Dr. Robert Reitano and the John Hancock Mutual Life Insurance Company for supplying the data.) The prior values, $\boldsymbol{\theta}^P$, are based on a graduation from an earlier mortality study. The policies in this study were similar to those policies comprising the current data set.

The basic data, together with four graduations under assumption (*I*) and four graduations under assumption (*IC*), are listed in Tables 1 and 2. The values of α and w corresponding to these eight graduations are displayed in Table 3. Plots of the prior values, maximum likelihood estimates, and graduations are displayed in Figures 1 to 8. For each assumption, the four graduations were obtained by varying the value of m. The values used for m were not necessarily intended to represent practical choices, but rather to demonstrate the range of results that is possible.

TABLE 1

DATA AND GRADUATED VALUES OF θ UNDER ASSUMPTION (I)

					$\theta^B(I)$			
Age	d	e	θ^P	θ^M	$m = 1$	$m = 5$	$m = 25$	$m = 10^{10}$
35.......	3	1771.5	0.0012308	0.00169	0.00098	0.00091	0.00088	0.00093
36.......	1	2126.5	0.0012808	0.00047	0.00103	0.00095	0.00091	0.00093
37.......	3	2743.5	0.0013609	0.00109	0.00111	0.00103	0.00098	0.00093
38.......	2	2766.0	0.0014811	0.00072	0.00122	0.00113	0.00105	0.00093
39.......	2	2463.0	0.0016213	0.00081	0.00137	0.00128	0.00118	0.00093
40.......	4	2368.0	0.0017816	0.00169	0.00158	0.00154	0.00153	0.00169
41.......	4	2310.0	0.0019519	0.00173	0.00179	0.00179	0.00179	0.00173
42.......	7	2306.5	0.0021423	0.00303	0.00204	0.00210	0.00215	0.00223
43.......	5	2059.5	0.0023628	0.00243	0.00229	0.00231	0.00229	0.00223
44.......	2	1917.0	0.0026134	0.00104	0.00256	0.00254	0.00243	0.00223
45.......	8	1931.0	0.0028942	0.00414	0.00298	0.00320	0.00346	0.00412
46.......	13	1746.5	0.0031951	0.00744	0.00335	0.00360	0.00383	0.00412
47.......	8	1580.0	0.0035362	0.00506	0.00360	0.00377	0.00392	0.00412
48.......	2	1580.0	0.0039377	0.00127	0.00385	0.00392	0.00400	0.00412
49.......	7	1467.5	0.0044097	0.00477	0.00421	0.00416	0.00414	0.00412
50.......	4	1516.0	0.0049422	0.00264	0.00457	0.00439	0.00427	0.00412
51.......	7	1371.5	0.0054850	0.00510	0.00503	0.00472	0.00447	0.00412
52.......	4	1343.0	0.0060382	0.00298	0.00548	0.00503	0.00464	0.00412
53.......	4	1304.0	0.0066017	0.00307	0.00608	0.00552	0.00495	0.00412
54.......	11	1232.5	0.0072663	0.00892	0.00716	0.00744	0.00795	0.00892
55.......	11	1204.5	0.0080523	0.00913	0.00825	0.00866	0.00905	0.00913
56.......	13	1113.5	0.0090710	0.01167	0.00962	0.01016	0.01053	0.01116
57.......	12	1048.0	0.0101210	0.01145	0.01075	0.01116	0.01131	0.01116
58.......	12	1155.0	0.0111823	0.01039	0.01184	0.01213	0.01205	0.01116
59.......	19	1018.5	0.0122548	0.01865	0.01308	0.01360	0.01410	0.01526
60.......	12	945.0	0.0133386	0.01270	0.01397	0.01428	0.01455	0.01526
61.......	16	853.0	0.0145047	0.01876	0.01497	0.01512	0.01521	0.01526
62.......	12	750.0	0.0158753	0.01600	0.01594	0.01579	0.01562	0.01526
63.......	6	693.0	0.0174514	0.00866	0.01701	0.01649	0.01603	0.01526
64.......	10	594.0	0.0192848	0.01684	0.01870	0.01807	0.01752	0.01684

d is the observed number of deaths.
e is the number of years of observation at the indicated age; see definition before (2.2).
θ^P is the prior value of θ.
θ^M is the maximum likelihood estimate of θ.
The corresponding mortality rate is defined in (1.2).

TABLE 2

DATA AND GRADUATED VALUES OF θ UNDER ASSUMPTION (*IC*)

Age	d	e	θ^P	θ^M	$\theta^B(IC)$ $m = 1$	$m = 50$	$m = 250$	$m = 10^{10}$
35.......	3	1771.5	0.0012308	0.00169	0.00098	0.00090	0.00091	0.00099
36.......	1	2126.5	0.0012808	0.00047	0.00104	0.00094	0.00093	0.00099
37.......	3	2743.5	0.0013609	0.00109	0.00113	0.00103	0.00099	0.00099
38.......	2	2766.0	0.0014811	0.00072	0.00127	0.00119	0.00116	0.00099
39.......	2	2463.0	0.0016213	0.00081	0.00143	0.00139	0.00136	0.00128
40.......	4	2368.0	0.0017816	0.00169	0.00162	0.00161	0.00161	0.00157
41.......	4	2310.0	0.0019519	0.00173	0.00181	0.00185	0.00186	0.00187
42.......	7	2306.5	0.0021423	0.00303	0.00203	0.00210	0.00213	0.00216
43.......	5	2059.5	0.0023628	0.00243	0.00227	0.00237	0.00242	0.00246
44.......	2	1917.0	0.0026134	0.00104	0.00255	0.00266	0.00271	0.00275
45.......	8	1931.0	0.0028942	0.00414	0.00285	0.00297	0.00302	0.00305
46.......	13	1746.5	0.0031951	0.00744	0.00317	0.00330	0.00333	0.00334
47.......	8	1580.0	0.0035362	0.00506	0.00353	0.00364	0.00366	0.00364
48.......	2	1580.0	0.0039377	0.00127	0.00394	0.00400	0.00399	0.00393
49.......	7	1467.5	0.0044097	0.00477	0.00442	0.00439	0.00435	0.00423
50.......	4	1516.0	0.0049422	0.00264	0.00495	0.00484	0.00473	0.00452
51.......	7	1371.5	0.0054850	0.00510	0.00550	0.00529	0.00513	0.00481
52.......	4	1343.0	0.0060382	0.00298	0.00606	0.00576	0.00553	0.00511
53.......	4	1304.0	0.0066017	0.00307	0.00663	0.00624	0.00595	0.00617
54.......	11	1232.5	0.0072663	0.00892	0.00731	0.00711	0.00699	0.00731
55.......	11	1204.5	0.0080523	0.00913	0.00812	0.00811	0.00810	0.00845
56.......	13	1113.5	0.0090710	0.01167	0.00916	0.00921	0.00925	0.00958
57.......	12	1048.0	0.0101210	0.01145	0.01024	0.01035	0.01043	0.01072
58.......	12	1155.0	0.0111823	0.01039	0.01132	0.01149	0.01161	0.01186
59.......	19	1018.5	0.0122548	0.01865	0.01241	0.01264	0.01280	0.01299
60.......	12	945.0	0.0133386	0.01270	0.01352	0.01381	0.01399	0.01413
61.......	16	853.0	0.0145047	0.01876	0.01470	0.01502	0.01522	0.01527
62.......	12	750.0	0.0158753	0.01600	0.01606	0.01631	0.01650	0.01640
63.......	6	693.0	0.0174514	0.00866	0.01761	0.01772	0.01784	0.01754
64.......	10	594.0	0.0192848	0.01684	0.01942	0.01935	0.01938	0.01868

d is the observed number of deaths.
e is the number of years of observation at the indicated age; see definition before (2.2).
θ^P is the prior value of θ.
θ^M is the maximum likelihood estimate of θ.
The corrresponding mortality rate is defined in (1.2).

TABLE 3

VALUES OF α AND w CORRESPONDING TO THE GRADUATIONS IN TABLES 1 AND 2

(*I*) m	α	w	(*IC*) m	α	w
1	2.311827652	0.28	1	2.332941843	0.18
5	1.467399490	0.35	50	1.131267399	0.21
25	1.188084363	0.42	250	1.056737850	0.26
10^{10}	1.000002728	0.55	10^{10}	1.000002760	0.30

An inspection of Figures 1 to 8 verifies the expected result: Decreasing m draws $\boldsymbol{\theta}^B(\cdot)$ closer to $\boldsymbol{\theta}^P$ and therefore increases smoothness. Although no examples are reported with $m < 1$, if small enough values were chosen for m, we would not be able to distinguish between $\boldsymbol{\theta}^B(\cdot)$ and $\boldsymbol{\theta}^P$.

Increasing m decreases the influence that the prior values, $\boldsymbol{\theta}^P$, have on the graduation. If we have little confidence in the prior values, then a large value should be used for m. Setting $m = 10^{10}$ allows the prior values to be completely overwhelmed by the data. To verify this, θ_i^P was replaced by $\theta_i^P + .01, i = 1, \ldots, 30$, and $\boldsymbol{\theta}^B(I)$ and $\boldsymbol{\theta}^B(IC)$ were recomputed with $m = 10^{10}$. The resulting graduations were identical to those reported in Tables 1 and 2 for $m = 10^{10}$. This indicates that the graduations in Figures 4 and 8 are effectively independent of the prior values. Although taking m large enough will eliminate the influence of $\boldsymbol{\theta}^P$ on the graduation, it will not eliminate the influence of the restriction. It must always be true that $\boldsymbol{\theta}^B(I) \in R_I$ and $\boldsymbol{\theta}^B(IC) \in R_{IC}$. (The graduations with $m = 10^{10}$ appear to violate this requirement, but that is because the data in the tables are limited to five-place accuracy.)

Suppose a graduation previously had been done for ages 34 and below, and we wanted the graduation for ages 35 through 64 to join it in a smooth manner. By "smooth" we may simply mean that $\theta_1^B(\cdot) > \theta_0^G$, where $\theta_0^G = -\ln(1 - q_{34}^G)$ and q_{34}^G is the graduated value of q_{34}. This may be accomplished by a respecification of our model. Under restriction (I) let

$$\theta_0 = \phi_0$$

$$\theta_1 = \phi_0 + \phi_1$$

$$\vdots$$

$$\theta_k = \phi_0 + \phi_1 + \ldots + \phi_k,$$

where ϕ_0 is a fixed quantity specified by the graduator and would normally be set equal to θ_0^G. The prior on $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_k)$ is still $g(\boldsymbol{\phi}|\mathbf{a},\mathbf{r})$, so the posterior is again given by (3.6), but with $\theta_j = \phi_0 + \ldots + \phi_j$.

Correspondingly, the Bayes estimate of θ_j is $\theta_j^B(I) = \phi_0 + \sum_{i=1}^{j} \phi_i^B(I)$, where $[\phi_1^B(I), \ldots, \phi_k^B(I)]$ is the posterior mode. The method for selecting **a** and **r** may be used just as before with the one exception that $\phi_1^P = \theta_1^P - \phi_0$ rather that θ_1^P.

Under restriction (IC) the reparameterization is

$$\begin{aligned}
\theta_0 &= \psi_0 \\
\theta_1 &= \psi_0 + \psi_1 \\
\theta_2 &= \psi_0 + \psi_1 + \psi_2 \\
\theta_3 &= \psi_0 + \psi_1 + 2\psi_2 + \psi_3 \\
&\;\;\vdots \\
\theta_k &= \psi_0 + \psi_1 + (k-1)\,\psi_2 + (k-2)\,\psi_3 + \ldots + \psi_k,
\end{aligned}$$

where ψ_0 denotes the value specified for θ_0. The posterior is given by (3.9) except

$$\begin{aligned}
\theta_j &= \psi_0 + \psi_1, & j &= 1 \\
&= \psi_0 + \psi_1 + \sum_{i=2}^{j} (j - i + 1)\psi_i, & j &= 2, \ldots, k,
\end{aligned}$$

and the Bayes estimate is

$$\begin{aligned}
\theta_j^B(IC) &= \psi_0 + \psi_1^B(IC), & j &= 1 \\
&= \psi_0 + \psi_1^B(IC) + \sum_{i=2}^{j} (j - i + 1)\,\psi_i^B(IC), & & \\
& & j &= 2, \ldots, k,
\end{aligned}$$

where $(\psi_1^B(IC), \ldots, \psi_k^B(IC))$ is the posterior mode. The previously described method for choosing **a** and **r** is applicable, but with ψ_1^P defined as $\theta_1^P - \psi_0$.

In order to demonstrate this technique and the more general method of selecting **a** and **r**, one additional graduation was computed. Assumption (I) was used with $\theta_0 = 0.00119$, $n = 2$, $k_1 = 24$, $k_2 = 6$, $m_1 = 30$, and $m_2 = 23$ (the lower bound for m_2 was 22.45). For this graduation $w = 0.38$, and the results are displayed in Figure 9. Specifying $\theta_1 > 0.00119$ caused an increase in the graduated values at the lower end of the age scale, and picking m_2 close to its lower bound drew the graduated values closer to $\boldsymbol{\theta}^P$ at the upper end of the age scale.

REFERENCES

1. BARLOW, R.E., BARTHOLOMEW, D.J., BREMNER, J.M., AND BRUNK, H.D. *Statistical Inference Under Order Restrictions*. New York: John Wiley & Sons, Inc., 1972.
2. BROFFITT, J.D. "Maximum Likelihood Alternatives to Actuarial Estimators of Mortality Rates," *TSA* XXXVI (1984): 77–122.
3. BROFFITT, J.D. "A Bayes Estimator for Ordered Parameters and Isotonic Bayesian Graduation," *Scandinavian Actuarial Journal* (1984): 231–47.
4. BROFFITT, J.D. "Isotonic Bayesian Graduation with an Additive Prior," *Advances in the Statistical Sciences*. Vol. 6, *Actuarial Science*, edited by I.B. MacNeill and G.J. Umphrey, 19–40. Boston: D. Reidel Publishing Co., 1986.
5. BROFFITT, J.D. "Restricted Bayes Estimators for Binomial Parameters," *Probability and Bayesian Statistics,* edited by R. Viertl. New York: Plenum Press, 1987, 61–72.
6. CHAN, L.K., AND PANJER, H.H. "A Statistical Approach to Graduation by Mathematical Formula," *Insurance: Mathematics & Economics* 2 (1983): 33–47.
7. DENT, W.T., AND ROBERTSON, T.J. "Consistent Estimation of Concave Nondecreasing Functions," unpublished manuscript, 1973.
8. HICKMAN, J.C., AND MILLER, R.B. "Notes on Bayesian Graduation," *TSA* XXIX (1977): 1–21.
9. HILDRETH, C. "Point Estimates of Ordinates of Concave Functions," *Journal of the American Statistical Association* 49 (1954): 598–619.
10. HOGG, R.V., AND CRAIG, A.T. *Introduction to Mathematical Statistics*, 4th ed. New York: Macmillan, 1978.
11. KIMELDORF, G.S., AND JONES, D.A. "Bayesian Graduation," *TSA* XIX (1967): 66–112.
12. SMITH, A.M.F. "A Bayesian Note on Reliability Growth During a Development Testing Program," *IEEE Transactions on Reliability* R-26 (1977): 346–47.
13. STEELMAN, J.M. "Statistical Approaches to Mortality Estimation," M.S. thesis, Department of Statistics, University of Iowa, Iowa City, 1968.

APPENDIX

Let $f_i(\mathbf{x}) = \frac{\partial}{\partial x_i} \ln \text{post}(\mathbf{x})$, $i = 1, \ldots, k$, where post(·) may represent either the posterior density of $\boldsymbol{\phi}$ under assumption (I) or the posterior density of $\boldsymbol{\psi}$ under (IC). The goal is to find $\mathbf{x}$ so that

$$f_i(\mathbf{x}) = 0, i = 1, \ldots, k. \tag{A.1}$$

Although the equations (A.1) differ slightly under the two restrictions, the iterative technique is the same, so in the following discussion it will not be necessary to distinguish between the restrictions (I) and (IC).

For $a_i > 1$ and $x_i > 0$, $i = 1, \ldots, k$, we obtain from either (3.7) or (3.10) the results

$$\frac{\partial}{\partial x_i} f_i(\mathbf{x}) < 0,$$

$$\frac{\partial^2}{\partial^2 x_i} f_i(\mathbf{x}) > 0,$$

$$f_i(\mathbf{x}) > 0 \text{ for } x_i \text{ sufficiently close to } 0,$$

and

$$f_i(\mathbf{x}) < 0 \text{ for } x_i \text{ sufficiently large.}$$

Thus for fixed $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k$, $f_i(\mathbf{x})$ is a decreasing convex function of x_i and there is a unique solution for x_i in $f_i(\mathbf{x}) = 0$. Although this does not prove that the system (A.1) has a unique solution, there have been no multiple solutions or convergence problems detected in practice.

Because the system (A.1) is well-behaved, it was possible to design the iterative procedure so that it does not require matrix manipulations. The starting value of the interation process is the prior value $\boldsymbol{\phi}^P$ or $\boldsymbol{\psi}^P$ (depending on which restriction is used) denoted by $\mathbf{x}^{(0)} = (x_1^{(0)}, \ldots, x_k^{(0)})$. Successively, for $i = 1, \ldots, k$, a new value is computed for x_i. This new value is labeled $x_i^{(1)}$, and after calculating $x_1^{(1)}, \ldots, x_i^{(1)}$, the current solution for (A.1) is denoted as

$$\mathbf{x}_i^{(0)} = (x_1^{(1)}, \ldots, x_i^{(1)}, x_{i+1}^{(0)}, \ldots, x_k^{(0)}).$$

The first iteration is completed after determining $x_k^{(1)}$. At this point the current solution is denoted by

$$\mathbf{x}^{(1)} = \mathbf{x}_0^{(1)} = \mathbf{x}_k^{(0)} = (x_1^{(1)}, \ldots, x_k^{(1)}).$$

In general, after completing $j+1$ full iterations, the current solution is denoted by any of $\mathbf{x}^{(j+1)}$, $\mathbf{x}_0^{(j+1)}$, or $\mathbf{x}_k^{(j)}$.

The second iteration is completed by successively computing new values for x_i for $i = 1, \ldots k$. This process is repeated until enough iterations have been done to obtain convergence. After computing a new value for x_i in the $(j+1)$th iteration, the current solution is

$$\mathbf{x}_i^{(j)} = (x_1^{(j+1)}, \ldots, x_i^{(j+1)}, x_{i+1}^{(j)}, \ldots, x_k^{(j)}),$$

and after computing $x_k^{(j+1)}$ the current solution is denoted by

$$\mathbf{x}^{(j+1)} = \mathbf{x}_0^{(j+1)} = \mathbf{x}_k^{(j)} = (x_1^{(j+1)}, \ldots, x_k^{(j+1)}).$$

The value $x_i^{(j+1)}$ is calculated by using Newton's method, that is,

$$x_i^{(j+1)} = x_i^{(j)} - \frac{f_i(\mathbf{x}_{i-1}^{(j)})}{f'_i(\mathbf{x}_{i-1}^{(j)})}, \tag{A.2}$$

where $f'_i(\mathbf{x}) = \dfrac{\partial}{\partial x_i} f_i(\mathbf{x})$. If (A.2) produces a nonpositive value, then $x_i^{(j+1)}$ is defined by $x_i^{(j+1)} = .5x_i^{(j)}$.

At the completion of each iteration, convergence is checked by comparing the percentage change in each x_i with a specified tolerance. That is, the process is stopped after iteration j if

$$\max_{1 \le i \le k} [100|x_i^{(j-1)} - x_i^{(j)}|/x_i^{(j-1)}] \le \text{tolerance}.$$

For the examples in section V, tolerance = 0.01, so that a change of .01 percent or less in each x_i is required to stop. The resulting number of iterations required for the graduations in Figure 1 to 9 were, respectively, 13, 22, 28, 67, 17, 114, 206, 643, and 28.

FIGURE 1

Plots of θ^M, θ^P, and $\theta^B(I)$ for $m = 1$.

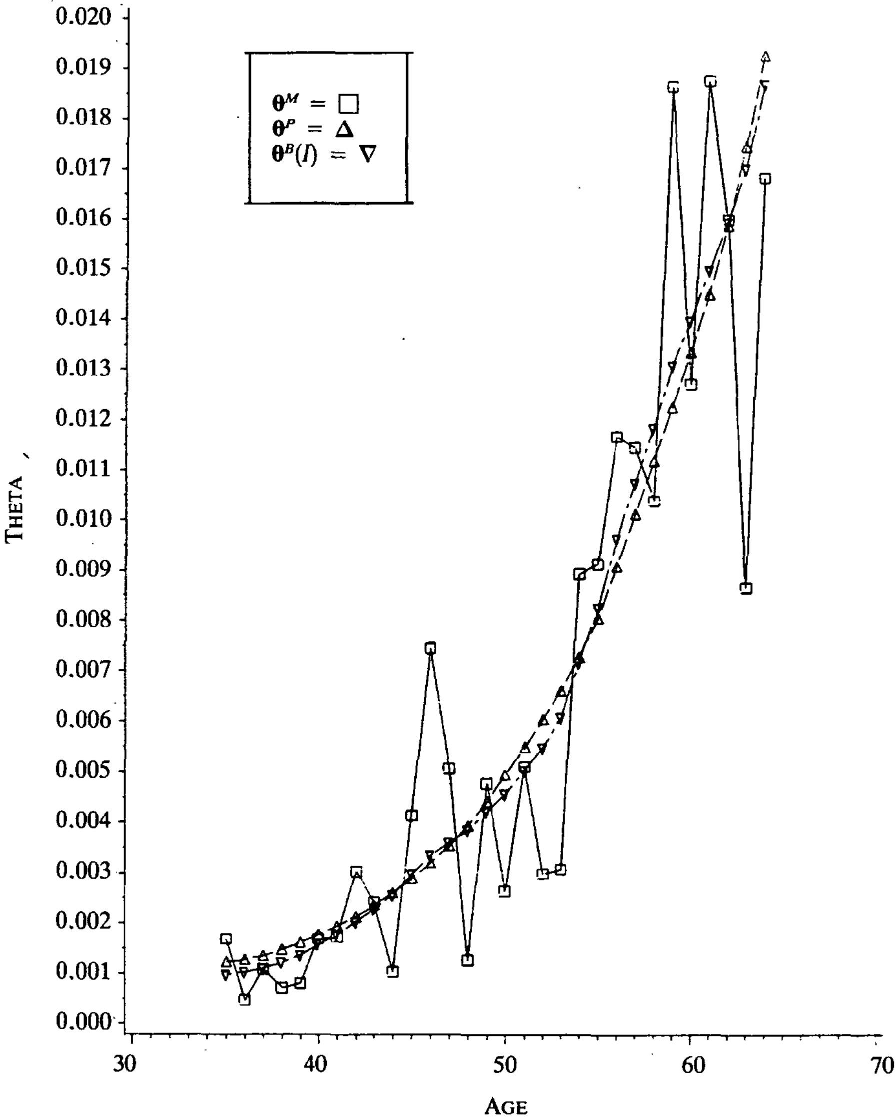


FIGURE 2

Plots of θ^M, θ^P, and $\theta^B(l)$ for $m = 5$

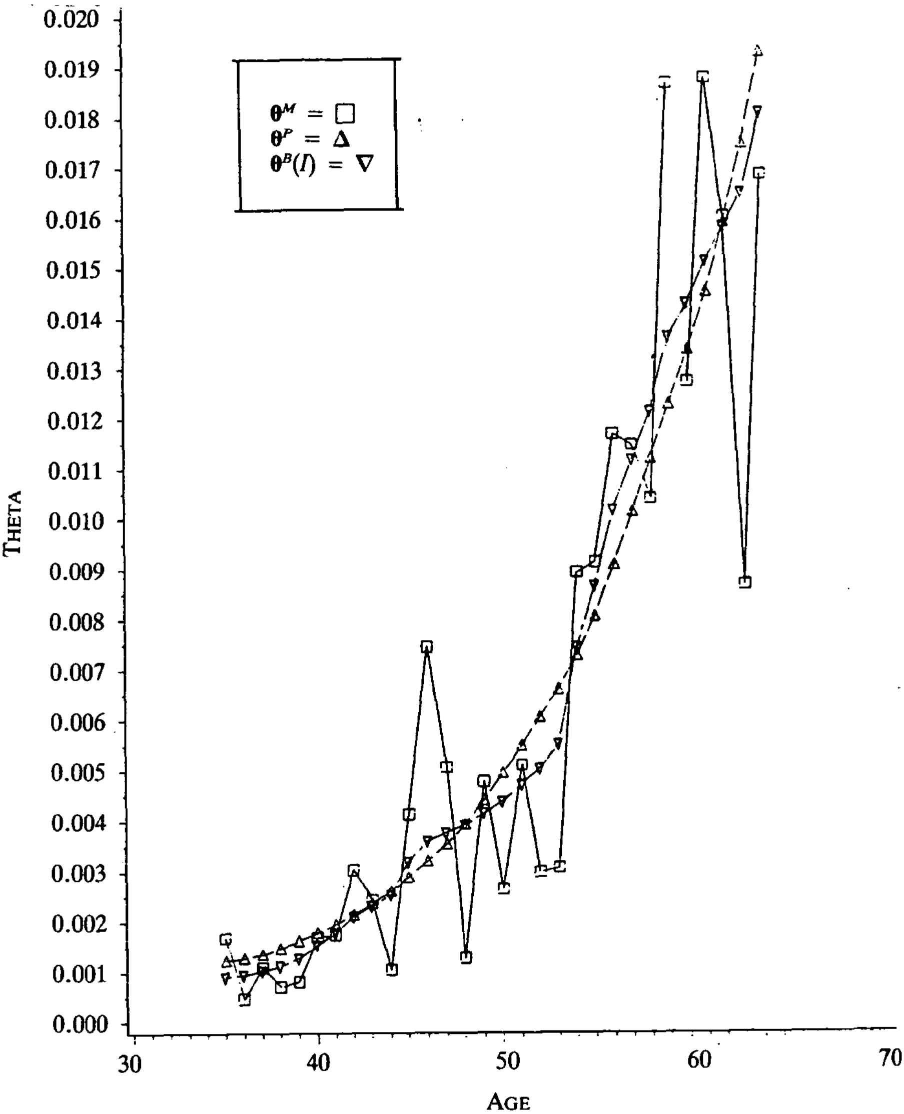


FIGURE 3

Plots of θ^M, θ^P, and $\theta^B(I)$ for $m = 25$

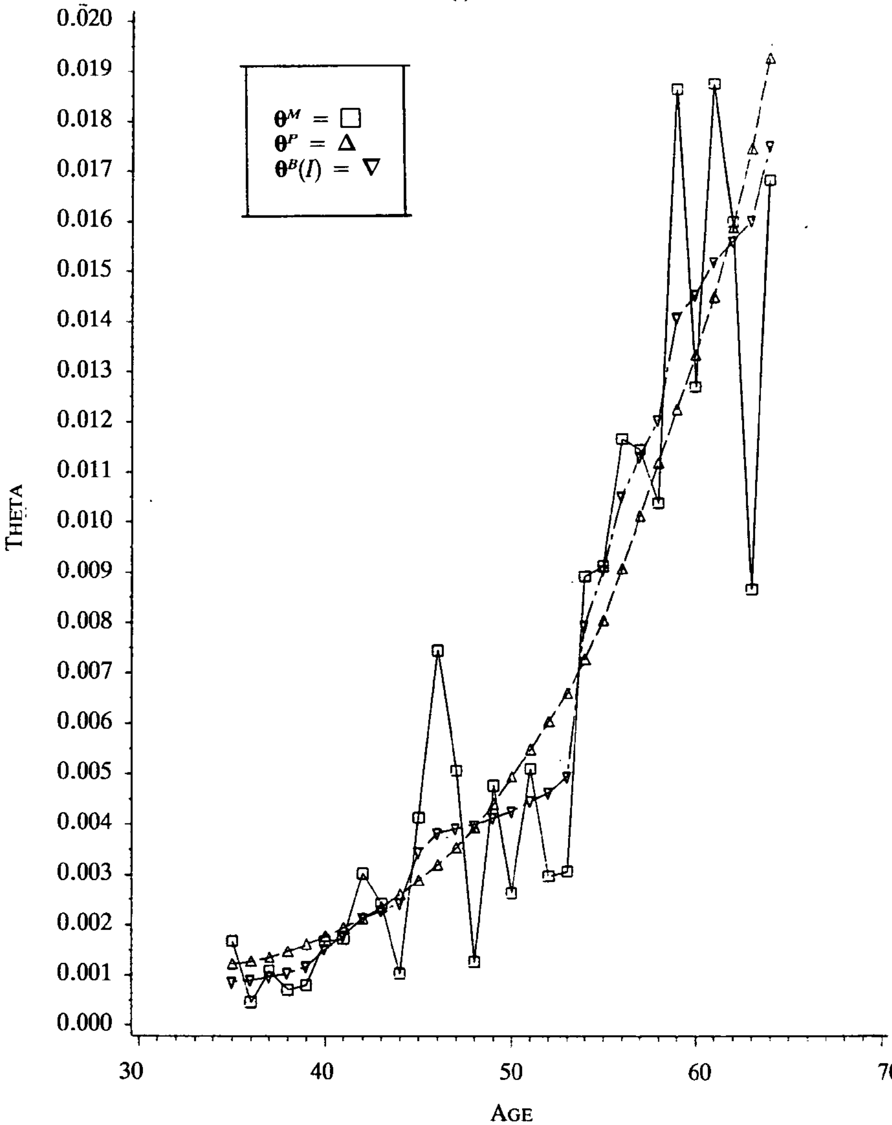


FIGURE 4

Plots of θ^M, θ^P, and $\theta^B(I)$ for $m = 10^{10}$

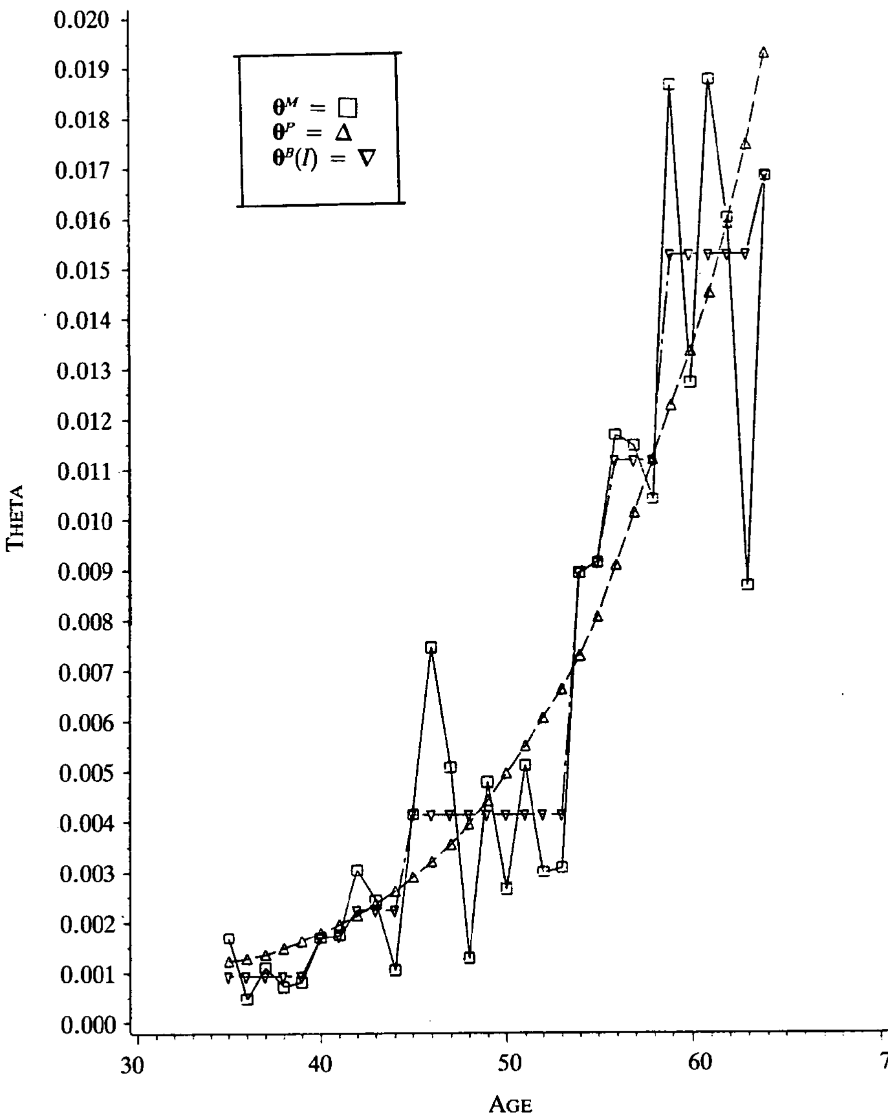


FIGURE 5

Plots of θ^M, θ^P, and $\theta^B(IC)$ for $m = 1$

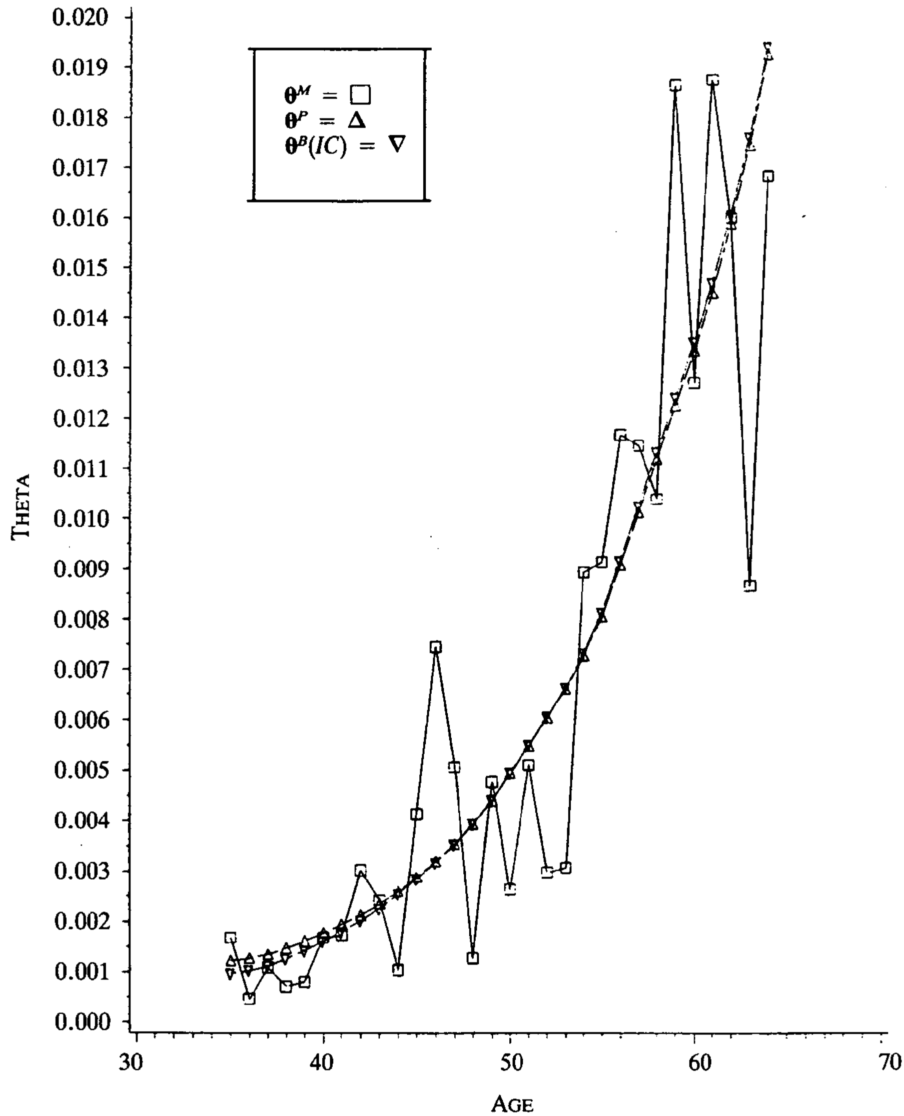


FIGURE 6

Plots of θ^M, θ^P, and $\theta^B(IC)$ for $m = 50$

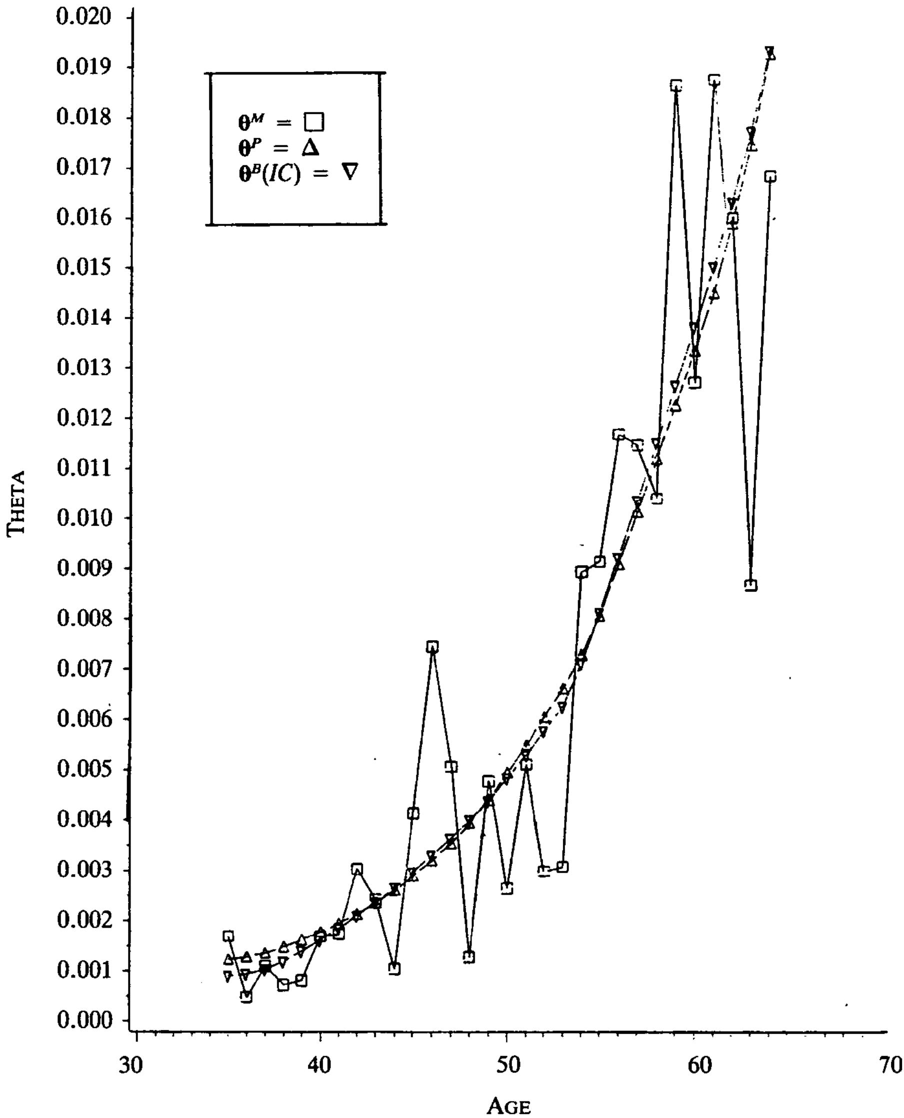


FIGURE 7

Plots of θ^M, θ^P, and $\theta^B(IC)$ for $m = 250$

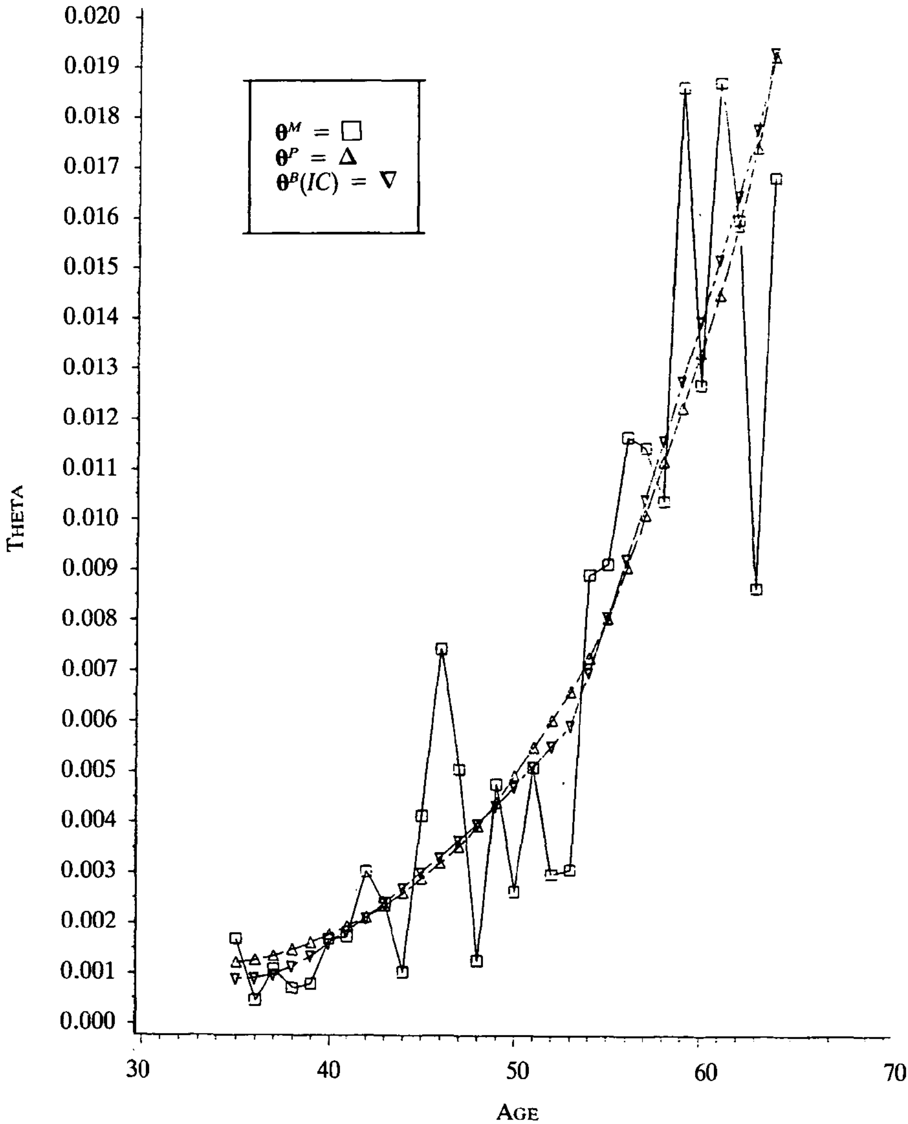


FIGURE 8

Plots of θ^M, θ^P, and $\theta^B(IC)$ for $m = 10^{10}$

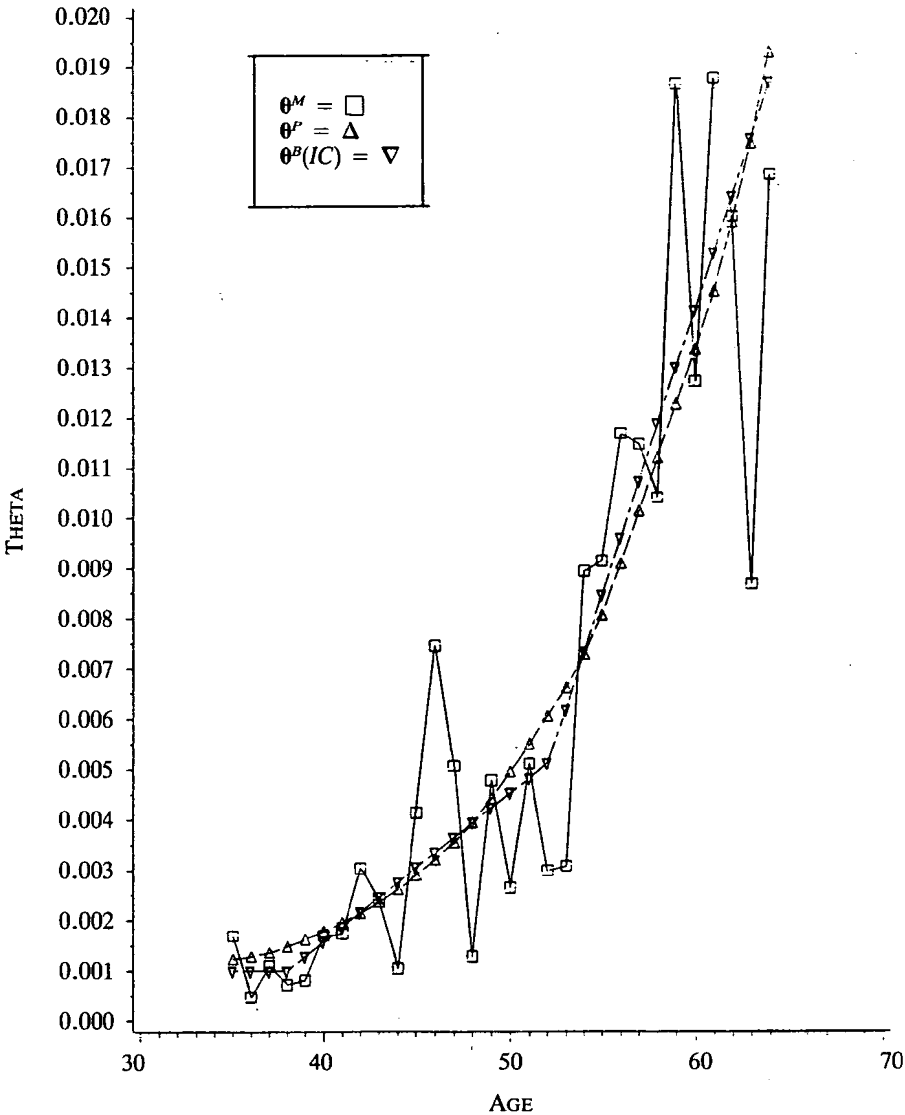
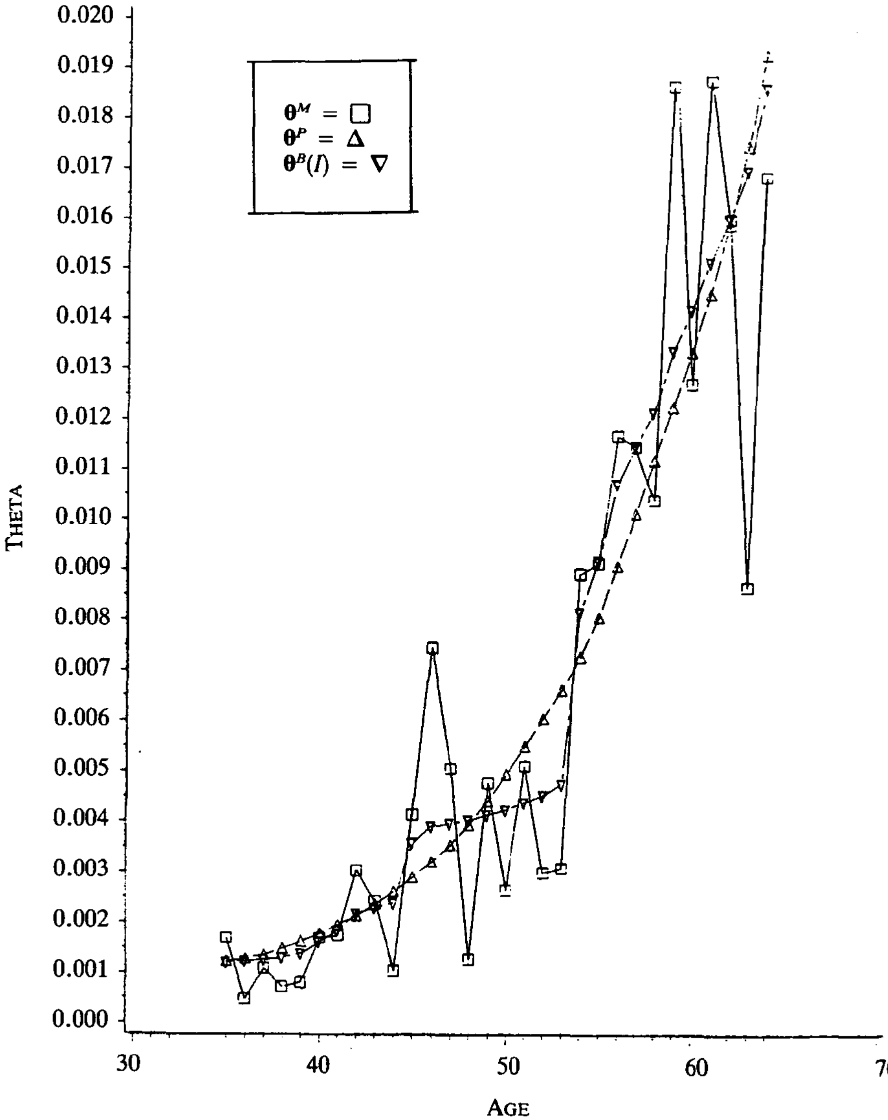


FIGURE 9

Plots of θ^M, θ^P, and $\theta^B(I)$ for $k_1 = 24$, $k_2 = 6$, $\theta_0 = 0.00119$, $m_1 = 30$, and $m_2 = 23$



DISCUSSION OF PRECEDING PAPER

THOMAS N. HERZOG:

Professor Broffitt deserves thanks for considering an interesting and important actuarial problem, as well as providing us with a lucid exposition of his work.

I just have a few questions/comments that I would like to raise to extend the discussion.

In many practical insurance problems, the result should be the construction of a predictive distribution (see Herzog*) of a set of random variables given the observed data. For example, in the calculation of net premium rates, the result should be the predictive distribution $P(X_1, X_2, \ldots, X_n |$ observed data), where n is the number of policies insured and X_i is the net gain or loss on the i-th policy. So, the net premium rate for year $t+1$ could be a function of the results observed during year t. Thus, the posterior distribution of the graduated values is not the final product, but rather an intermediate result. Consequently, I think that for many practical actuarial problems (for example, the calculation of premium rates, reserve requirements, and so on), point estimates of mortality parameter are, by themselves, of little use.

If the actuary does not desire to use the full predictive distribution approach but is familiar with the process generating the mortality data, then he or she may often do just as well to use a simple graphical procedure to graduate mortality rates, rather than a more sophisticated mathematical one. In this simplistic fashion, the actuary can ensure that the mortality rates are monotonically increasing with age.

For 30-year term FHA-insured single-family mortgages, the annual claim termination rates decrease monotonically after the third policy year. By using a simple transformation, one can apply "increasing graduation" methods to mortgage guarantee insurance problems as well as life insurance problems.

(AUTHOR'S REVIEW OF DISCUSSION)

JAMES D. BROFFITT:

I am grateful to Dr. Herzog for his kind remarks and for providing a discussion of my paper.

I cannot disagree that the predictive distribution of future net gains or losses is of primary interest. However, pragmatically I believe we are a long

*HERZOG, T.N., *An Introduction to Bayesian Credibility*, Part 4 Study Note. New York: Casualty Actuarial Society, 1985.

way from using such a distribution as the sole basis for setting premiums and reserves in life insurance. It is difficult to imagine that a life actuary would ever exclude a life table from his or her toolbox.

Graphical graduation is a simple technique that allows the imposition of restrictions (such as increasing rates) on the graduated values. Over the years a good deal of energy has been devoted to developing "mathematical" graduation, and some of these methods may produce results that are inferior to graphical graduation. However, this does not mean that research on mathematical graduation should cease or that such graduations should not be used because graphical graduation is simpler. There are some disadvantages to graphical graduation, and once programmed, a mathematical graduation is simpler to produce than a graphical one. This is the case with my technique. After the data are entered into a file, a few keystrokes and a few seconds' wait provide a table of the graduated values.

As Dr. Herzog points out, the techniques presented can easily be applied to other restrictions. For example, to estimate decreasing rates, one merely needs to index the data in reverse order so that the first subscript refers to the smallest rate and the last subscript denotes the largest rate.

In closing, I wish to express my appreciation to the Committee on Papers for making many helpful suggestions and, in particular, to J.C. McKenzie Smith, who checked my calculations.