Refinements of Whittaker Graduation Through the Minimization of Bayes Risk

James D. Broffitt*
Statistics & Actuarial Science
University of Iowa
Iowa City, Iowa 52242
e:mail: broffitt@uiowa.edu

October 1992

Abstract

Actuaries generally view the estimation of a sequence of mortality rates as a two-stage process: (1) obtain crude rates at each age (initial estimates), and (2) adjust these initial estimates to form a smooth sequence (final estimates). The second stage is called graduation and the final estimates are called graduated values. Standard Whittaker graduation determines graduated values by minimizing the quantity $F + hS$, where F and S are the usual lack-of-fit and lack-of-smoothness measures, and h is a tuning constant which allows the actuary to adjust the emphasis placed on smoothness relative to fit. In practice the value of h is generally determined by trial and error rather than by sound statistical reasoning.

In this paper the accuracy of a graduation process is measured by Bayes risk. Let $L(\mathbf{v}, \boldsymbol{\theta})$ be the loss function (a measure of discrepancy between the graduation $\mathbf{v}$ and the true mortality rates $\boldsymbol{\theta}$). Then the risk function is $R_{\mathbf{v}}(\boldsymbol{\theta}) = E[L(\mathbf{v}, \boldsymbol{\theta}) \mid \boldsymbol{\theta}]$, and the Bayes risk is the average of $R_{\mathbf{v}}(\boldsymbol{\theta})$ over the prior distribution of $\boldsymbol{\theta}$. For Whittaker graduation the Bayes risk may be evaluated after specifying the first two prior moments of $\boldsymbol{\theta}$ rather than the entire prior distribution. With the objective of minimizing Bayes risk, we first replace S by a measure of lack-of-smoothness that is 'better' than the standard one, and then determine the optimal value of h. These results are demonstrated on real data sets.

*My thanks to The Actuarial Education and Research Fund for providing partial financial support for this project.

1 Introduction

Let $\boldsymbol{\theta}$ be the $n \times 1$ vector of true mortality rates for n consecutive ages or age intervals of interest, and let $\mathbf{u}$ be the corresponding vector of n initial estimates (ungraduated values) of $\boldsymbol{\theta}$. Standard Whittaker graduation determines graduated values, $\mathbf{v}$, which minimize

$$F + hS \equiv (\mathbf{v} - \mathbf{u})'\mathbf{W}(\mathbf{v} - \mathbf{u}) + h\mathbf{v}'\mathbf{K}'\mathbf{K}\mathbf{v}, \tag{1}$$

where $\mathbf{K}$ is an $n - z \times n$ matrix such that $\mathbf{Kv}$ is the vector of zth differences of $v_1, \cdots, v_n$, and $\mathbf{W}$ is a diagonal matrix whose diagonal elements are specified weights for the various ages or age groups. We assume the diagonal elements of $\mathbf{W}$ are positive so that $\mathbf{W}$ is a positive definite matrix. It is well known that (1) is minimized when

$$\mathbf{v} = (\mathbf{W} + h\mathbf{K}'\mathbf{K})^{-1}\mathbf{W}\mathbf{u}. \tag{2}$$

For example see London [8].

The value of h is usually determined by trial and error. If $\mathbf{v}$ does not display enough smoothness, h is increased, and conversely, if the fit of $\mathbf{v}$ to $\mathbf{u}$ is unacceptable, then h is decreased. The value of h is continually adjusted until the results 'look' right. Although we may never be able to completely eliminate this type of tinkering, the practice of good statistics mandates that the estimator (i.e., formula for $\mathbf{v}$) be determined before the data is examined. Aside from the argument of statistical validity, it makes good practical sense to have an objective method of (at least approximately) determining h. This research was initiated with this goal in mind. Before attacking the problem of determining h, however, we will first suggest replacing $S = \mathbf{v}'\mathbf{K}'\mathbf{K}\mathbf{v}$ by a measure of lack-of-smoothness which provides graduated values that are 'better' than those given by (2). The replacement of S will be taken up in section 2, and the choice of h is explored in section 3. The remaining section deals with the application of these techniques to mortality data. Proofs have been relegated to an appendix in order to enhance readability.

1.1 Elementary Example

As an aid to understanding the results given in the following sections, we present an elementary illustration of comparing estimators. Suppose θ is a uni-dimensional unknown parameter in the interval (0,1), and let u be an unbiased estimator of θ with variance $B(\theta)$. That is, $E(u \mid \theta) = \theta$ and $Var(u \mid \theta) = B(\theta)$. Let h be a constant and define the estimator $v(h) = h \cdot u$. We now have an infinite set of estimators of θ to consider: $\{v(h) : -\infty < h < \infty\}$, and we desire to determine which one is 'best.' (Notice that many other estimators could be considered since *any* function of u could be viewed as an estimator of θ.)

How do we compare estimators? This is often done by examining mean squared error (MSE), which is the risk function associated with a squared error loss function.

That is, $L[v(h),\theta] = [v(h)-\theta]^2$, and $MSE = E\{L[v(h),\theta] \mid \theta\}$. Of course we prefer the estimator with the smallest MSE.

$$\begin{aligned} MSE_{v(h)}(\theta) &= E\{[v(h)-\theta]^2 \mid \theta\} \\ &= Var[v(h) \mid \theta] + \{E[v(h) \mid \theta] - \theta\}^2 \\ &= h^2 B(\theta) + (h-1)^2\theta^2 \end{aligned} \tag{3}$$

Now suppose $B(\theta) = \theta(1-\theta)/9$ and consider the two estimators $v(1.0)$ and $v(0.8)$. From (3) we have $MSE_{v(1.0)}(\theta) = \theta(1-\theta)/9$ and $MSE_{v(0.8)}(\theta) = (0.64\theta - 0.28\theta^2)/9$. These functions are displayed in Figure 1. For $\theta = 0.5$ the two estimators have the same MSE. If $\theta < 0.5$, then $v\,(0.8)$ is better, and if $\theta > 0.5$, $v\,(1.0)$ is better. Since we do not know the value of θ, there is not a clear choice between $v(1.0)$ and $v(0.8)$. If, however, we are certain that $\theta < 0.5$, then there is no contest as $v(0.8)$ is the clear winner.

Without additional information about θ the comparison of MSE does not provide a unique best estimator. This problem arises because we are comparing *functions* of θ ($MSE_{v(0.8)}(\theta)$ and $MSE_{v(1.0)}(\theta)$) and one is not always smaller than the other. A decisive comparison could be made if we reduced each MSE function to a single number. A natural choice is to average $MSE(\theta)$ over the values of θ according to a weight function which assigns higher weights to the more likely values of θ. Let $p(\theta)$ be a weight function which is scaled so that $\int_0^1 p(\theta)d\theta = 1$. That is, $p(\theta)$ is the prior *pdf* of θ. The average of the MSE is called the Bayes risk and we will denote it by BR. Then

$$\begin{aligned} BR &= \int_0^1 MSE(\theta)p(\theta)d\theta \\ &= E\{[v(h)-\theta]^2\}, \end{aligned}$$

where the last expectation is w.r.t. the joint *pdf* of u and θ.

In our example the MSEs are linear in θ and θ^2, and therefore to compute their Bayes risks we do not need to know the entire prior distribution but only its first two moments. Suppose, for example, that $E(\theta) = 0.1$ and $E(\theta^2) = 0.04$, then $BR_{v(0.8)} = 0.0528/9 < BR_{v(1.0)} = 0.06/9$, and therefore $v(0.8)$ is better than $v(1.0)$. In general, from (3) (and still assuming $B(\theta) = \theta(1-\theta)/9$, $E(\theta) = 0.1$, and $E(\theta^2) = 0.04$),

$$BR_{v(h)} = 0.04 - 0.08h + (0.42/9)h^2,$$

which is minimized when $h = 6/7$. Thus, for the assumed prior moments and model for u, $v(6/7)$ is the best estimator of θ in the class $\{v(h) : -\infty < h < \infty\}$, w.r.t. a squared error loss function.

1.2 Model Specification

The vector of parameters to be estimated, $\boldsymbol{\theta}$, is assumed to be an unobservable realization of a random vector $\boldsymbol{\theta}$ (for convenience our notation will not distinguish between random variables and their realizations) with mean vector $E(\boldsymbol{\theta}) = \mathbf{m}$ and dispersion matrix $D(\boldsymbol{\theta}) = \mathbf{A} = [Cov(\theta_i, \theta_j)]$. To use the techniques developed in this paper it is not necessary to specify the entire prior distribution of $\boldsymbol{\theta}$, rather we need only specify the prior parameters $\mathbf{m}$ and $\mathbf{A}$. This is generally difficult to do, as is often the case in Bayesian analysis. In the graduation setting this important problem is considered in the papers by Kimeldorf and Jones [6] and Hickman and Miller [5] and their corresponding discussions. In section 4 an empirical Bayes method will be used to estimate $\mathbf{A}$. The graduator, however, is at liberty to specify $\mathbf{m}$ and $\mathbf{A}$ by any method that is deemed appropriate.

We assume the initial estimators are unbiased, i.e., $E(\mathbf{u} \mid \boldsymbol{\theta}) = \boldsymbol{\theta}$, with dispersion matrix $D(\mathbf{u} \mid \boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\theta})$, and $E[\mathbf{B}(\boldsymbol{\theta})] = \mathbf{B}$, where the last expectation is over the distribution of $\boldsymbol{\theta}$.

As a measure of discrepancy between $\mathbf{v}$ and $\boldsymbol{\theta}$ we use a quadratic loss function, i.e., $L(\mathbf{v}, \boldsymbol{\theta}) = (\mathbf{v} - \boldsymbol{\theta})'\mathbf{C}(\mathbf{v} - \boldsymbol{\theta})$, where $\mathbf{C}$ is a specified (symmetric) positive definite matrix. The corresponding Bayes risk is

$$BR_{\mathbf{v}} = E[(\mathbf{v} - \boldsymbol{\theta})'\mathbf{C}(\mathbf{v} - \boldsymbol{\theta})]. \tag{4}$$

In (4) the expectation is w.r.t. the joint distribution of $\mathbf{v}$ and $\boldsymbol{\theta}$. The comparison of graduation techniques is then based on a comparison of their Bayes risks. Under this criteria the graduation process which produces graduations denoted by $\mathbf{v}^{(1)}$ is better than the process which produces graduations denoted by $\mathbf{v}^{(2)}$, if and only if

$$E[(\mathbf{v}^{(1)} - \boldsymbol{\theta})'\mathbf{C}(\mathbf{v}^{(1)} - \boldsymbol{\theta})] < E[(\mathbf{v}^{(2)} - \boldsymbol{\theta})'\mathbf{C}(\mathbf{v}^{(2)} - \boldsymbol{\theta})]. \tag{5}$$

Notice that when we say $\mathbf{v}^{(1)}$ is better than $\mathbf{v}^{(2)}$, we do not mean that for a particular set of data $L(\mathbf{v}^{(1)}, \boldsymbol{\theta}) < L(\mathbf{v}^{(2)}, \boldsymbol{\theta})$. Rather, we mean that relation (5) holds, i.e., *on the average* $\mathbf{v}^{(1)}$ will have a smaller loss than $\mathbf{v}^{(2)}$. When considering a class of estimators (graduation techniques) we will select the one that minimizes the Bayes risk. In the sequel, the minimum Bayes risk criteria will be used twice: first to replace the 'smoothness' measure, $\mathbf{v}'\mathbf{K}'\mathbf{K}\mathbf{v}$ by a better one, and second to select the best value of h.

A recent paper by Taylor [12] places Whittaker graduation in a Bayesian context. He assumes that, asymptotically, the elements of $\mathbf{u}$ are independent normal variates with mean vector $\boldsymbol{\theta}$, and $\mathbf{W} = [D(\mathbf{u} \mid \boldsymbol{\theta})]^{-1}$. The prior distribution is expressed through $\Delta^z\theta_i$, which, for $i = 1, \cdots, n - z$, are assumed to be *iid* $N(0, \tau^2)$. He then concludes that the posterior mean of $\boldsymbol{\theta}$ is the Whittaker graduation with $h = 1/\tau^2$. Through a similar argument Carlin and Klugman [4] reach essentially the same conclusion. They continue the analysis by assuming the parameters $\boldsymbol{\theta}$ and τ^2 are unknown and use a hierarchical Bayesian model to estimate $\boldsymbol{\theta}$. It is logical that h should be inversely related to τ^2 since, if τ^2 is large, then the prior information is not precise and we should give a larger weight to fit.

Our goal is not to give a Bayesian justification for Whittaker graduation. Rather it is to use Bayesian analysis to improve upon the practice of Whittaker graduation. The results we develop in sections 2 and 3 are nonparametric in nature since they do not depend on distributional assumptions; however, we will make use of the multivariate normal distribution in section 4.

2 Modification of S

For convenience let $\mathbf{H} = h\mathbf{K}'\mathbf{K}$, and extend (1) by replacing $\mathbf{v}$ with $\mathbf{v} - \mathbf{x}$ in the smoothness term S. The graduated values are then obtained by minimizing

$$(\mathbf{v} - \mathbf{u})'\mathbf{W}(\mathbf{v} - \mathbf{u}) + (\mathbf{v} - \mathbf{x})'\mathbf{H}(\mathbf{v} - \mathbf{x})$$

w.r.t. $\mathbf{v}$ and, since $\mathbf{W} + \mathbf{H}$ is positive definite, are given by

$$\mathbf{v} = (\mathbf{W} + \mathbf{H})^{-1}(\mathbf{W}\mathbf{u} + \mathbf{H}\mathbf{x}). \tag{6}$$

Setting $\mathbf{x} = \mathbf{0}$ gives the standard Whittaker graduation; however, we are free to choose $\mathbf{x}$ to provide optimal results. The vector $\mathbf{x}$ will be selected to minimize the Bayes risk of (6).

Theorem 1 *Denote the Bayes risk of graduation (6) by* $BR(\mathbf{x}) = E[(\mathbf{v} - \boldsymbol{\theta})'\mathbf{C}(\mathbf{v} - \boldsymbol{\theta})]$. *Let* $\mathbf{L} = \mathbf{W} + \mathbf{H}$ *and* $\mathbf{G} = \mathbf{L}^{-1}\mathbf{C}\mathbf{L}^{-1}$. *Then*

$$BR(\mathbf{x}) = tr(\mathbf{WGWB} + \mathbf{HGHA}) + (\mathbf{x} - \mathbf{m})'\mathbf{HGH}(\mathbf{x} - \mathbf{m}),$$

where $tr(\mathbf{Q})$ *denotes the traee (sum of the diagonal elements) of the square matrix* $\mathbf{Q}$.

Since $\mathbf{G}$ is positive definite, $(\mathbf{x} - \mathbf{m})'\mathbf{HGH}(\mathbf{x} - \mathbf{m}) \geq 0$. Thus $BR(\mathbf{x})$ will be minimized if $\mathbf{x}$ is selected so that $(\mathbf{x} - \mathbf{m})'\mathbf{HGH}(\mathbf{x} - \mathbf{m}) = 0$. Let $y(\cdot)$ be *any* polynomial of degree $z - 1$ or less, and let $\mathbf{y} = (y(1), \cdots, y(n))'$. Since $\mathbf{K}\mathbf{y} = \mathbf{0}$, it is clear that $BR(\mathbf{x})$ is minimized if $\mathbf{x} = \mathbf{m} + \mathbf{y}$. Thus the optimal choice of $\mathbf{x}$ is not unique; it is $\mathbf{m}$ plus any $n \times 1$ vector whose elements lie on a polynomial of degree $z - 1$ or less. For any such $\mathbf{x}$, graduation (6) becomes

$$\mathbf{v} = (\mathbf{W} + \mathbf{H})^{-1}(\mathbf{W}\mathbf{u} + \mathbf{H}\mathbf{m}) \tag{7}$$

(which is invariant w.r.t. $\mathbf{y}$), and its Bayes risk is $BR(\mathbf{m} + \mathbf{y}) = BR(\mathbf{m}) = tr(\mathbf{WGWB} + \mathbf{HGHA})$. Graduation (7) is an appealing result since the graduated values are a linear combination of the initial estimates and the prior mean.

The standard Whittaker graduation (2) has Bayes risk $BR(\mathbf{0})$ which exceeds $BR(\mathbf{m})$ by $\mathbf{m}'\mathbf{HGH}\mathbf{m}$. This is strictly positive unless the elements of $\mathbf{m}$ lie on a polynomial of degree $z - 1$ or less, in which case $\mathbf{m}'\mathbf{HGH}\mathbf{m} = 0$. In this latter case graduations (2) and (7) are identical to each other (since $\mathbf{K}\mathbf{m} = \mathbf{0}$). However, in the more common case where $\mathbf{Km} \neq \mathbf{0}$, graduation (7) is better than (2) since $BR(\mathbf{m}) < BR(\mathbf{0})$.

In summary, we suggest that the smoothness term $\mathbf{v}'\mathbf{K}'\mathbf{K}\mathbf{v}$ be replaced by $(\mathbf{v}-\mathbf{m})'\mathbf{K}'\mathbf{K}(\mathbf{v}-\mathbf{m})$. Doing so will provide the graduations given in (7).

For either graduations (2) or (7) the standard for ultimate smoothness is attained by letting $h = \infty$, since this requires the lack-of-smoothness term to be 0. The graduation $\mathbf{v}$ will then be the vector which minimizes $F = (\mathbf{v}-\mathbf{u})'\mathbf{W}(\mathbf{v}-\mathbf{u})$ subject to the lack-of-smoothness term being 0. For standard Whittaker graduation this implies that the elements of $\mathbf{v}$ will lie on the polynomial of degree $z-1$ that is fit to $\mathbf{u}$ by weighted least squares. For our modification, $\mathbf{v}$ would equal $\mathbf{m}$ plus a vector whose elements lie on the polynomial of degree $z-1$ that is fit to $\mathbf{u}-\mathbf{m}$ by weighted least squares.

It is instructive to examine such graduations. In Example 1, which is described in section 4, the ages range from 15 to 100. For this example, Figures 2a through 2c contain plots of $\mathbf{u}$ and $\mathbf{v}$ obtained by setting $h = \infty$ in standard Whittaker graduation for $z = 1, 2, 3$, and 4. (The plots have been split into three figures with different scales in order to provide a better display of the information.) Being low degree polynomials these graduations are quite smooth, however it is clear that for $z = 1$ and 2 the fit is particularly poor. Figures 3a through 3c provide the corresponding plots for the modified graduations. The curve labeled 'prior' is a plot of $\mathbf{m}$. With $h = \infty$ varying z has much less effect in modified Whittaker than it does in standard Whittaker. The modified graduations do not appear to be as smooth as the standard ones; however, this is due to the 'roughness' displayed by $\mathbf{m}$. In particular the graduations in Figure 3c are nearly parallel to $\mathbf{m}$. Is this wrong? No, not if $\mathbf{m}$ accurately represents our prior opinion of the true mortality rates. We should then be pleased that $\mathbf{v}$ displays the same pattern as $\mathbf{m}$. However, if we believe that the $\mathbf{m}$ plotted in Figures 3a through 3c is not an adequate representation of our prior opinion, then $\mathbf{m}$ should be modified before the graduation process continues.

In order to compute (7) we must know the prior mean of $\boldsymbol{\theta}$, i.e., $\mathbf{m}$. Suppose we misspecify this mean. That is, $E(\boldsymbol{\theta}) = \mathbf{m}$, but we erroneously assume $E(\boldsymbol{\theta}) = \mathbf{s}$, where $\mathbf{s} \neq \mathbf{m}$. The resulting graduation is

$$\mathbf{v} = (\mathbf{W}+\mathbf{H})^{-1}(\mathbf{Wu}+\mathbf{Hs}), \tag{8}$$

which has Bayes risk, $BR(\mathbf{s})$. Now

$$BR(\mathbf{0}) - BR(\mathbf{s}) = \mathbf{m}'\mathbf{HGHm} - (\mathbf{s}-\mathbf{m})'\mathbf{HGH}(\mathbf{s}-\mathbf{m}).$$

Graduation (8) will still be better than (2) if this difference is positive. For example, suppose we either under-specified each element of $\mathbf{m}$ by 75%, or we over-specified each element of $\mathbf{m}$ by 75%. Then $\mathbf{s} = \frac{1}{4}\mathbf{m}$ or $\mathbf{s} = \frac{7}{4}\mathbf{m}$, and in either case $BR(\mathbf{0}) - BR(\mathbf{s}) = \frac{7}{16}\mathbf{m}'\mathbf{HGHm} \geq 0$, with strict inequality holding except in the unlikely event that the elements of $\mathbf{m}$ lie on a polynomial of degree $z-1$ or less.

We may view $\mathbf{m}'\mathbf{HGHm}$ as a measure of distance between $\mathbf{0}$ and $\mathbf{m}$, in the metric specified by $\mathbf{HGH}$. Also, $(\mathbf{s}-\mathbf{m})'\mathbf{HGH}(\mathbf{s}-\mathbf{m})$ is the corresponding measure of distance between $\mathbf{s}$ and $\mathbf{m}$. The conclusion is that we do not need to know $\mathbf{m}$ exactly, or even approximately, to produce graduations that are superior to standard Whittaker graduations (2). As long as our assignment of $\mathbf{s}$ is 'closer' to $\mathbf{m}$, than $\mathbf{0}$ is to $\mathbf{m}$, graduations given by (8) will have smaller Bayes risk than standard Whittaker graduations.

Our modified graduation (7) is obtained by minimizing the objective function

$$(\mathbf{v}-\mathbf{u})'\mathbf{W}(\mathbf{v}-\mathbf{u})+h(\mathbf{v}-\mathbf{m})'\mathbf{K}'\mathbf{K}(\mathbf{v}-\mathbf{m}).$$

This has the flavor of the Bayesian graduation developed by Kimeldorf and Jones [6]. They assumed multivariate normal distributions for the data and the prior. The corresponding graduated values are obtained by minimizing

$$(\mathbf{v}-\mathbf{u})'[\mathbf{B}(\mathbf{m})]^{-1}(\mathbf{v}-\mathbf{u})+(\mathbf{v}-\mathbf{m})'\mathbf{A}^{-1}(\mathbf{v}-\mathbf{m}).$$

Thus the modified graduation (7) may be viewed as a compromise between traditional Whittaker and Bayesian graduation.

3 Selection of h

For the remainder of this paper we will concentrate on the estimator

$$\mathbf{v}=(\mathbf{W}+h\mathbf{K}'\mathbf{K})^{-1}(\mathbf{Wu}+h\mathbf{K}'\mathbf{Km}). \tag{9}$$

Since we are interested in finding h to minimize the Bayes risk of (9), we will use the notation $BR(h)$. From Theorem 1,

$$BR(h)=tr(\mathbf{WGWB}+\mathbf{HGHA}).$$

Let $\mathbf{W}^{1/2}$ be the diagonal matrix whose diagonal elements are the square roots of the corresponding diagonal elements of $\mathbf{W}$, and $\mathbf{W}^{-1/2}$ be the inverse of $\mathbf{W}^{1/2}$. Since $\mathbf{W}^{-1/2}\mathbf{K}'\mathbf{K}\mathbf{W}^{-1/2}$ is positive semidefinite with rank $n-z$, $\mathbf{W}^{-1/2}\mathbf{K}'\mathbf{K}\mathbf{W}^{-1/2}$ has $n-z$ positive characteristic roots and z characteristic roots equal to 0. Denote these roots by

$$\lambda_1\geq\cdots\geq\lambda_{n-z}>\lambda_{n-z+1}=\cdots=\lambda_n=0.$$

Also let

$$\underset{n\times n}{\mathbf{P}}=[p_{ij}]=[\mathbf{p}_1,\cdots,\mathbf{p}_n]$$

be an orthogonal matrix (i.e., $\mathbf{P}^{-1}=\mathbf{P}'$) which diagonalizes $\mathbf{W}^{-1/2}\mathbf{K}'\mathbf{K}\mathbf{W}^{-1/2}$, so that

$$\mathbf{W}^{-1/2}\mathbf{K}'\mathbf{K}\mathbf{W}^{-1/2}=\mathbf{P}\mathbf{D}_\lambda\mathbf{P}' \tag{10}$$

where $\mathbf{D}_\lambda$ is the $n \times n$ diagonal matrix whose kth diagonal element is λ_k. Finally, let

$$\begin{array}{lll} \alpha_{ij} & \text{be the } i,j\text{th element of} & \mathbf{P'W^{1/2}AW^{1/2}P}, \\ \beta_{ij} & \text{be the } i,j\text{th element of} & \mathbf{P'W^{1/2}BW^{1/2}P}, \text{ and} \\ \gamma_{ij} & \text{be the } i,j\text{th element of} & \mathbf{P'W^{-1/2}CW^{-1/2}P}. \end{array}$$

Notice that

$$\begin{aligned} \alpha_{ij} &= \mathbf{p}_i'\mathbf{W}^{1/2}\mathbf{A}\mathbf{W}^{1/2}\mathbf{p}_j \\ &= \sum_{r=1}^{n}\sum_{s=1}^{n} a_{rs}(w_r w_s)^{1/2} p_{ri} p_{sj}. \end{aligned} \tag{11}$$

Correspondingly,

$$\beta_{ij} = \sum_{r=1}^{n}\sum_{s=1}^{n} b_{rs}(w_r w_s)^{1/2} p_{ri} p_{sj}, \tag{12}$$

and

$$\gamma_{ij} = \sum_{r=1}^{n}\sum_{s=1}^{n} c_{rs}(w_r w_s)^{-1/2} p_{ri} p_{sj}. \tag{13}$$

The matrices $\mathbf{P}$ and $\mathbf{D}_\lambda$ can be obtained from a computer package that does matrix computations. Then (11), (12), and (13) may be used to calculate α_{ij}, β_{ij}, and γ_{ij}.

Theorem 2 *The Bayes risk of (9) is*

$$BR(h) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\gamma_{ij}(\beta_{ij} + \alpha_{ij}\lambda_i\lambda_j h^2)}{(1+\lambda_i h)(1+\lambda_j h)}.$$

In practice a natural choice for the matrix $\mathbf{C}$ is the diagonal weight matrix $\mathbf{W}$. This is reasonable since if we assign a 'large' weight to a particular age when computing $\mathbf{v}$, it follows that the loss function should incur a 'large' penalty if the difference between the true rate and graduated rate is 'large' for this age. When $\mathbf{W}$ is substituted for $\mathbf{C}$, $\mathbf{P'W^{-1/2}CW^{-1/2}P} = \mathbf{I}$, and thus $\gamma_{ij} = \delta_{ij}$ (Kronecker delta). The Bayes risk may then be simplified further as in the following corollary to Theorem 2.

Corollary 3 *If $\mathbf{C} = \mathbf{W}$, then the Bayes risk of (9) is*

$$BR(h) = \sum_{i=1}^{n-z} \frac{\beta_i + \alpha_i\lambda_i^2 h^2}{(1+\lambda_i h)^2} + \sum_{i=n-z+1}^{n} \beta_i, \tag{14}$$

where $\beta_i = \beta_{ii}$ and $\alpha_i = \alpha_{ii}$.

Let

$$f_i(h) = \frac{\beta_i + \alpha_i\lambda_i^2 h^2}{(1+\lambda_i h)^2}.$$

When $\mathbf{C} = \mathbf{W}$, $BR(h)$ is minimized when h is chosen to minimize $\sum_{i=1}^{n-z} f_i(h)$. The following properties of $f_i(h)$ are easily verified:

1. $f_i(0) = \beta_i$

2. $f_i(+\infty) = \alpha_i$

3. $f_i'(h) < 0 \Leftrightarrow h < \beta_i/(\alpha_i\lambda_i)$

From these properties it follows that:

1. $BR(0) = \sum_{i=1}^{n} \beta_i$

2. $BR(+\infty) = \sum_{i=1}^{n-z} \alpha_i + \sum_{i=n-z+1}^{n} \beta_i$

3. $BR(h)$ achieves a minimum for some h in the interval $[\min_{i\leq n-z} \beta_i/(\alpha_i\lambda_i), \max_{i\leq n-z} \beta_i/(\alpha_i\lambda_i)]$

The optimal value of h may be found by using numerical techniques to minimize (14).

4 Application

In this section we address some practical issues associated with determining h and the corresponding graduated values. Of special interest are methods of assigning the prior parameters $\mathbf{m}$ and $\mathbf{A}$, and the matrix $\mathbf{B}$. There is not a single correct solution. Depending on the particular application and the amount of prior information available, there may be a variety of reasonable choices for $\mathbf{m}$, $\mathbf{A}$, and $\mathbf{B}$. As previously stated, the assignment of prior parameters is one of the most difficult problems in Bayesian analysis and a good deal of care should be devoted to it. In our examples we will use an empirical Bayes method to specify $\mathbf{A}$. The assignment of $\mathbf{m}$ is not as difficult since there is usually a table of standard rates which form the basis for $\mathbf{m}$. These rates may be graduated values from an earlier study of similar data. Some adjusting of these tabulated values will be necessary in case they do not completely satisfy our prior opinion.

When the initial estimates are based on frequency counts of the numbers of lives exposed and the numbers of deaths, it is customary to assume the numbers of deaths are independent binomial random variables. The independence assumption may not always be desirable; a good discussion on this point was given by Robinson [9]. Under the binomial model the *arcsin* transformation suggested by Hickman and Miller [5] provides a good approximation to the normal distribution in addition to stabilizing the variance. That is, if e_i and d_i are the exposure and death counts for the *ith* age, $u_i = d_i/e_i$, and $g(u) = \arcsin(\sqrt{u})$, then

$$g(u_i) \dot{\sim} N(g(\theta_i), \frac{1}{4e_i}), \tag{15}$$

for large values of e_i. One advantage of this transformation is that the variance term in (15) does not depend on the unknown parameter θ_i.

We find the *arcsin* transformation desirable and we will use it throughout our analysis. Thus the first step in analyzing a problem is to compute $g(u_i)$ and $g(m_i)$ for $i = 1, \cdots, n$. The graduation is then done with these transformed values. Consequently the graduated values are considered estimates of $g(\theta_i)$, $i = 1, \cdots, n$, and the transformation must be reversed to obtain estimates of the true rates. That is, if $g(v_i)$ denotes the graduated value for the *ith* mortality rate based on the transformed data, then the estimate of θ_i is $(\sin[g(v_i)])^2$. In order to keep the notation from becoming too cluttered, we will still use the symbols $\mathbf{u}, \boldsymbol{\theta}, \mathbf{m}$, and $\mathbf{v}$, but they will represent vectors of the transformed values. Similarly, $\mathbf{A}$ and $\mathbf{B}$ will denote the dispersion matrices associated with the transformed variables.

Based on (15) and the independence assumption, we let $\mathbf{B}(\boldsymbol{\theta}) = \mathbf{B} = \sigma^2\mathbf{D}$ where $\mathbf{D}$ is a diagonal matrix with *ith* diagonal element equal to $1/(4e_i)$. There are two choices for σ^2 depending on whether the data is based on counts of lives or counts of dollars. In the former case $\sigma^2 = 1$, and in the latter σ^2 will be a parameter that must be assigned a value. This is the same approach used by Carlin and Klugman [4]. If the distribution of the number of dollars across policies were known, then σ^2 could be determined (see Klugman [7]); however, if this information is not available, σ^2 may be included as a parameter to be estimated with the empirical Bayes method.

Let $\mathbf{A} = (\frac{\tau^2}{4\bar{e}})\mathbf{R}$ where $\mathbf{R} = [\rho^{|i-j|}]$ is a correlation matrix. Then the variance of θ_i is $\tau^2/(4\bar{e})$ for each i and the correlation between θ_i and θ_j is $\rho^{|i-j|}$. The divisor $4\bar{e}$ is a scaling factor that puts τ^2 on approximately the same level as σ^2. This model for $\mathbf{A}$ introduces two additional parameters, τ^2 and ρ, that must be specified.

We assume that $\mathbf{m}$ is known, but σ^2, τ^2, and ρ may be unknown. The approach we use in the following examples estimates these parameters by an empirical Bayes method. Under the assumptions that

$$\mathbf{u} \mid \boldsymbol{\theta} \sim N(\boldsymbol{\theta}, \mathbf{B}) \text{ and } \boldsymbol{\theta} \sim N(\mathbf{m}, \mathbf{A}),$$

it follows that the marginal distribution of $\mathbf{u}$ is given by

$$\mathbf{u} \sim N(\mathbf{m}, \mathbf{B} + \mathbf{A}).$$

If $f(\mathbf{u})$ denotes the marginal *pdf* of $\mathbf{u}$, then, aside from an additive constant,

$$-2\ln f(\mathbf{u}) = \ln(|\sigma^2\mathbf{D} + (\frac{\tau^2}{4\bar{e}})\mathbf{R}|) + (\mathbf{u} - \mathbf{m})'(\sigma^2\mathbf{D} + (\frac{\tau^2}{4\bar{e}})\mathbf{R})^{-1}(\mathbf{u} - \mathbf{m}). \qquad (16)$$

For given $\mathbf{u}$, $f(\mathbf{u})$ may be viewed as a likelihood function for σ^2, τ^2, and ρ. The empirical Bayes estimates of σ^2, τ^2, and ρ are then found by maximizing $f(\mathbf{u})$, or equivalently, by minimizing (16). Of course in the cases where $\sigma^2 = 1$, (16) is minimized w.r.t. τ^2 and ρ only.

The final quantity that needs to be assigned is $\mathbf{W}$. It seems reasonable that the diagonal elements of $\mathbf{W}$ should be proportional to the corresponding exposures.

Accordingly, we let $w_{ii} = e_i/\bar{e}$. The scaling factor $\bar{e}$ is used to keep the values of h from becoming too large.

Example 1: The first example is a graduation of the data upon which the 1975-80 Ultimate Basic Table for male lives was constructed. This data is contained in The Society of Actuaries 1982 Reports [11]. Table 1 contains a listing of this data together with the results of our modified Whittaker graduations. The 1975-80 Basic Table was a replacement for the 1965-70 Basic Table which is found in The Society of Actuaries 1973 Reports [10]. The values given in the 1965-70 Basic table were used for $\mathbf{m}$. Smoothness of the resulting graduations could be increased by altering $\mathbf{m}$ to increase its smoothness prior to doing the graduations; however, we did not adjust these tabulated values. Whether or not such an adjustment is warranted is a matter of personal taste.

Since the data is in terms of dollars, the empirical Bayes method was used to estimate σ^2 together with τ^2 and ρ. These estimates are $\widehat{\sigma}^2 = 214,698$, $\widehat{\tau}^2 = 4,168,358$, and $\widehat{\rho} = 0.9975$. The optimal value of h was then determined for $z = 1,2,3$ and 4 by minimizing (14). The results are given in Table 2, and plots of the Bayes risks are displayed in Figure 4. Finally, for each z, Figures 5a through 5c show the resulting modified Whittaker graduations. The graduator is still faced with the decision of which z to use. This problem can also be solved with the minimum Bayes risk criterion. By comparing the Bayes risk at the optimal h for the various values of z (Table 2) we see that, over-all, the Bayes risk is minimized when $z = 1$.

It is of interest to compare these modified Whittaker graduations to the standard Whittaker graduation which was the basis for the 1975-80 Ultimate Basic Table. This graduation was done on the untransformed data with $\mathbf{W} = \mathbf{I}$, $z = 2$, and $h = 18$, and the resulting values of v_i are plotted in Figures 6a through 6c. The 1982 Reports states that "The smoothness of the graduated rates at ages 85 and higher, as measured by second differences, was deemed not acceptable and empirical adjustments were therefore made which produced not only a very smooth curve (constant second differences to the end of the table) but improved the fit." From an examination of Figure 6c smoothness and fit appear to be adequate; however, the graduated rates start to decrease at age 95, and this should be considered unacceptable. Since standard Whittaker graduation does not formally use prior information, it is difficult to avoid this undesirable result in cases where the u_is display a decreasing pattern.

This result should emphasize the value of a formal Bayesian procedure which blends the data with prior information. Broffitt [1] and [2] developed a Bayesian graduation method which *requires* the graduated values to satisfy certain restrictions such as, "increase with age." Carlin [3] showed how this type of graduation could be computed using Gibbs sampling. Carlin and Klugman [4] applied this technique as well as their hierarchical Bayesian Whittaker graduation to the data of this example.

To conclude this example we computed a Kimeldorf Jones Bayesian graduation using the same prior parameters as above, i.e., $\mathbf{m}$ and $\mathbf{A}$. This graduation is displayed in Figure 6a through 6c. Surprisingly, this appears to be a poor graduation as smoothness suffers badly. Decreasing the value of τ^2 should correct this problem.

Example 2: For variety a final example is included where the data is based on counts of lives. The data and resulting graduations are given in Table 3. The exposures are somewhat small and consequently the ungraduated values display a very jagged pattern—as seen in Figures 7a through 7c. In this problem we set $\sigma^2 = 1$, and then minimized (16) to find $\widehat{\tau}^2 = 0.3730754$ and $\widehat{\rho} = 0.7493$. For $z = 1, 2, 3$ and 4, Table 2 contains the optimal values of h along with the Bayes risks. Again, $z = 1$ gives the smallest over-all Bayes risk. The graduations are plotted in Figures 7a through 7c.

Some data sets may be such that the empirical Bayes estimate of ρ equals one. In this case the graduator may prefer to use a different method to specify the matrix **A**. However, it is of interest to investigate the implications of using $\rho = 1$ in our modified Whittaker graduation.

Let $\rho = 1$. Then $a_{ij} = k$, where $k = \tau/(4\bar{e})$, and from (11)

$$\alpha_i = \alpha_{ii} = k(\sum_{r=1}^{n} w_r^{1/2} p_{ri})^2.$$

Let $x_i = \sqrt{\alpha_i/k}$, $\mathbf{x}_1' = (x_1, \cdots, x_{n-z})$, $\mathbf{x}_2' = (x_{n-z+1}, \cdots, x_n)$, and $\mathbf{x}' = (\mathbf{x}_1', \mathbf{x}_2')$. Notice that $\mathbf{x}' = \mathbf{1}'\mathbf{W}^{1/2}\mathbf{P}$, where $\mathbf{1}' = (1, \cdots, 1)$. Finally, let

$$\mathbf{D}_\lambda = \begin{bmatrix} \underset{n-z\times n-z}{\mathbf{D}_1} & \underset{n-z\times z}{\mathbf{0}} \\ \underset{z\times n-z}{\mathbf{0}} & \underset{z\times z}{\mathbf{0}} \end{bmatrix},$$

where $\mathbf{D}_1$ is a diagonal matrix whose diagonal elements are the positive characteristic roots of $\mathbf{W}^{-1/2}\mathbf{K}'\mathbf{K}\mathbf{W}^{-1/2}$, and $\mathbf{0}$ denotes a matrix of zeros. From (10)

$$\begin{aligned} \mathbf{1}'\mathbf{W}^{1/2}\mathbf{P}\mathbf{D}_\lambda\mathbf{P}'\mathbf{W}^{1/2}\mathbf{1} &= \mathbf{1}'\mathbf{K}'\mathbf{K}\mathbf{1} \\ \mathbf{x}'\mathbf{D}_\lambda\mathbf{x} &= 0 \\ \mathbf{x}_1'\mathbf{D}_1\mathbf{x}_1 &= 0 \end{aligned}$$

But since $\mathbf{D}_1$ is positive definite this implies $\mathbf{x}_1 = \mathbf{0}$ and therefore $\alpha_i = 0$ for $i = 1, \cdots, n-z$. Consequently, for $i = 1, \cdots, n-z$, $f_i(h) = \beta_i/(1 + \lambda_i h)^2$ decreases as h increases, and therefore $BR(h)$ is minimized when $h = \infty$.

References

[1] Broffitt, J.D. "A Bayes Estimator for Ordered Parameters and Isotonic Bayesian Graduation," *Scandinavian Actuarial Journal* (1984): 231-47.

[2] Broffitt, J.D. "Increasing and Increasing Convex Bayesian Graduation," *TSA* XL (1988): 115-148.

[3] Carlin, B.P. "A Simple Monte Carlo Approach to Bayesian Graduation," *TSA* (1993): To appear.

[4] Carlin, B.P. and Klugman, S.A. "Hierarchical Bayesian Graduation ," unpublished paper.

[5] Hickman, J.C. and Miller, R.B. "Notes on Bayesian Graduation," *TSA* XXIX (1977): 1-21.

[6] Kimeldorf, G.S. and Jones, D.A. "Bayesian Graduation," *TSA* XIX (1967): 66-112.

[7] Klugman, S.A. "On the Variance and Mean Squared Error of Decrement Estimators," *TSA* XXXIII (1981): 301-311.

[8] London, D. *Graduation: The Revision of Estimates.* Winsted and Abington, CT: ACTEX Publications, 1985.

[9] Robinson, J.M. Discussion of "Maximum Likelihood Alternatives to Actuarial Estimators of Mortality Rates," *TSA* XXXVI (1984): 125-138.

[10] Society of Actuaries. "1965-70 Basic Tables," *TSA 1973 Reports of Mortality and Morbidity Experience*, (1974): 199-223.

[11] Society of Actuaries. "1975-80 Basic Tables," *TSA 1982 Reports of Mortality and Morbidity Experience,* (1985): 55-81.

[12] Taylor, G. "A Bayesian Interpretation of Whittaker-Henderson Graduation," *Insurance: Mathematics & Economics* **11** (1992): 7-16.

A Appendix

We will make use of the following results:

- Let $\mathbf{A}$ and $\mathbf{B}$ be matrices of order $n \times m$ and $m \times n$, respectively. Then

$$\begin{aligned} tr(\mathbf{AB}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}b_{ji} && (17) \\ &= tr(\mathbf{BA}) && (18) \end{aligned}$$

- Let $\mathbf{x}$ be a random vector with mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$. Then

$$E(\mathbf{xx}') = \boldsymbol{\Sigma} + \boldsymbol{\mu\mu}'. \tag{19}$$

- Using the fact that $\mathbf{x}'\mathbf{Ay}$ is a scalar and (18),

$$\mathbf{x}'\mathbf{Ay} = tr(\mathbf{x}'\mathbf{Ay}) = tr[\mathbf{Ayx}']. \tag{20}$$

- Let $\mathbf{A}$ be a non-random matrix. Then from (20)

$$E(\mathbf{x}'\mathbf{Ay}) = tr[\mathbf{A}E(\mathbf{yx}')]. \tag{21}$$

A.1 Proof of Theorem 1

Since $\mathbf{v} = \mathbf{L}^{-1}(\mathbf{Wu} + \mathbf{Hx})$ and $\boldsymbol{\theta} = \mathbf{L}^{-1}\mathbf{L}\boldsymbol{\theta} = \mathbf{L}^{-1}(\mathbf{W}\boldsymbol{\theta} + \mathbf{H}\boldsymbol{\theta})$, it follows that

$$\mathbf{v} - \boldsymbol{\theta} = \mathbf{L}^{-1}\mathbf{W}(\mathbf{u} - \boldsymbol{\theta}) + \mathbf{L}^{-1}\mathbf{H}(\mathbf{x} - \boldsymbol{\theta}).$$

Using this together with $\mathbf{G} = \mathbf{L}^{-1}\mathbf{CL}^{-1}$, we obtain

$$\begin{aligned} (\mathbf{v} - \boldsymbol{\theta})'\mathbf{C}(\mathbf{v} - \boldsymbol{\theta}) &= (\mathbf{u} - \boldsymbol{\theta})'\mathbf{WGW}(\mathbf{u} - \boldsymbol{\theta}) \\ &+ 2(\mathbf{u} - \boldsymbol{\theta})'\mathbf{WGH}(\mathbf{x} - \boldsymbol{\theta}) \\ &+ (\mathbf{x} - \boldsymbol{\theta})'\mathbf{HGH}(\mathbf{x} - \boldsymbol{\theta}). \end{aligned}$$

Therefore by (21),

$$\begin{aligned} E[(\mathbf{v} - \boldsymbol{\theta})'\mathbf{C}(\mathbf{v} - \boldsymbol{\theta})] &= tr\mathbf{WGW}E[(\mathbf{u} - \boldsymbol{\theta})(\mathbf{u} - \boldsymbol{\theta})'] \\ &+ 2tr\mathbf{WGH}E[(\mathbf{x} - \boldsymbol{\theta})(\mathbf{u} - \boldsymbol{\theta})'] \\ &+ tr\mathbf{HGH}E[(\mathbf{x} - \boldsymbol{\theta})(\mathbf{x} - \boldsymbol{\theta})']. \end{aligned} \tag{22}$$

Now

$$E[(\mathbf{u}-\boldsymbol{\theta})(\mathbf{u}-\boldsymbol{\theta})'] = E\{E[(\mathbf{u}-\boldsymbol{\theta})(\mathbf{u}-\boldsymbol{\theta})' \mid \boldsymbol{\theta}]\} = E\{\mathbf{B}(\boldsymbol{\theta})\} = \mathbf{B},$$

$$E[(\mathbf{x}-\boldsymbol{\theta})(\mathbf{u}-\boldsymbol{\theta})'] = E\{E[(\mathbf{x}-\boldsymbol{\theta})(\mathbf{u}-\boldsymbol{\theta})' \mid \boldsymbol{\theta}]\} = E\{\mathbf{0}\} = \mathbf{0}$$

since $\mathbf{x}$ is non-random, and

$$E[(\mathbf{x}-\boldsymbol{\theta})(\mathbf{x}-\boldsymbol{\theta})'] = \mathbf{A} + (\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})', \text{ by (19)}.$$

When these three expressions are substituted into (22) we obtain

$$\begin{aligned} E[(\mathbf{v}-\boldsymbol{\theta})'\mathbf{C}(\mathbf{v}-\boldsymbol{\theta})] &= tr\mathbf{WGWB} \\ &+ tr\mathbf{HGH}[\mathbf{A} + (\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})']. \end{aligned}$$

Therefore, using (18),

$$E[(\mathbf{v}-\boldsymbol{\theta})'\mathbf{C}(\mathbf{v}-\boldsymbol{\theta})] = tr\mathbf{WGWB} + tr\mathbf{HGHA} + (\mathbf{x}-\mathbf{m})'\mathbf{HGH}(\mathbf{x}-\mathbf{m}),$$

which completes the proof.

A.2 Proof of Theorem 2

When $\mathbf{x} = \mathbf{m}$,

$$E[(\mathbf{v}-\boldsymbol{\theta})'\mathbf{C}(\mathbf{v}-\boldsymbol{\theta})] = tr\mathbf{WGWB} + tr\mathbf{HGHA} \tag{23}$$

where $\mathbf{G} = \mathbf{L}^{-1}\mathbf{C}\mathbf{L}^{-1}$ and $\mathbf{L} = \mathbf{W} + \mathbf{H} = \mathbf{W} + h\mathbf{K}'\mathbf{K}$. Making use of (10) we have

$$\begin{aligned} \mathbf{W}^{1/2}\mathbf{L}^{-1}\mathbf{W}^{1/2} &= \mathbf{W}^{1/2}(\mathbf{W}+h\mathbf{K}'\mathbf{K})^{-1}\mathbf{W}^{1/2} \\ &= (\mathbf{I}+h\mathbf{W}^{-1/2}\mathbf{K}'\mathbf{K}\mathbf{W}^{-1/2})^{-1} \\ &= (\mathbf{I}+h\mathbf{P}\mathbf{D}_\lambda\mathbf{P}')^{-1} \\ &= \mathbf{P}(\mathbf{I}+h\mathbf{D}_\lambda)^{-1}\mathbf{P}' \\ &= \mathbf{P}\mathbf{D}_{\frac{1}{1+\lambda h}}\mathbf{P}', \end{aligned} \tag{24}$$

where $\mathbf{D}_{\frac{1}{1+\lambda h}}$ denotes a diagonal matrix whose ith diagonal element is $\frac{1}{1+\lambda_i h}$.

We will consider each of the terms on the right hand side of (23).

$$\begin{aligned} tr\mathbf{WGWB} &= tr\mathbf{W}\mathbf{L}^{-1}\mathbf{C}\mathbf{L}^{-1}\mathbf{W}\mathbf{B} \\ &= tr\mathbf{W}^{1/2}\mathbf{W}^{1/2}\mathbf{L}^{-1}\mathbf{W}^{1/2}\mathbf{W}^{-1/2}\mathbf{C}\mathbf{W}^{-1/2}\mathbf{W}^{1/2}\mathbf{L}^{-1}\mathbf{W}^{1/2}\mathbf{W}^{1/2}\mathbf{B} \\ &= tr\mathbf{D}_{\frac{1}{1+\lambda h}}\mathbf{P}'\mathbf{W}^{-1/2}\mathbf{C}\mathbf{W}^{-1/2}\mathbf{P}\mathbf{D}_{\frac{1}{1+\lambda h}}\mathbf{P}'\mathbf{W}^{1/2}\mathbf{B}\mathbf{W}^{1/2}\mathbf{P} && (25) \\ &= tr\mathbf{D}_{\frac{1}{1+\lambda h}}[\gamma_{ij}]\mathbf{D}_{\frac{1}{1+\lambda h}}[\beta_{ij}] && (26) \end{aligned}$$

$$= tr\left[\frac{\gamma_{ij}}{1+\lambda_i h}\right]\left[\frac{\beta_{ij}}{1+\lambda_i h}\right] \tag{27}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\gamma_{ij}}{1+\lambda_i h}\cdot\frac{\beta_{ji}}{1+\lambda_j h} \tag{28}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\gamma_{ij}\beta_{ij}}{(1+\lambda_i h)(1+\lambda_j h)} \tag{29}$$

In the above development, (25) follows from (24) and (18), and (26) follows from (13) and (12). In (27) the brackets denote $n \times n$ matrices, and the contents of the brackets denote the $ijth$ elements of the corresponding matrices. Equation (28) follows from (17), and (29) follows since $[\beta_{ij}]$ is a symmetric matrix.

Notice that

$$\mathbf{W}^{-1/2}\mathbf{H}\mathbf{W}^{-1/2} = h\mathbf{W}^{-1/2}\mathbf{K}'\mathbf{K}\mathbf{W}^{-1/2} = h\mathbf{P}\mathbf{D}_\lambda\mathbf{P}', \tag{30}$$

and consider the second term on the right hand side of (23).

$$\begin{aligned}
tr\mathbf{HGHA} &= tr\mathbf{HL}^{-1}\mathbf{CL}^{-1}\mathbf{HA} \\
&= tr\mathbf{W}^{1/2}\mathbf{W}^{-1/2}\mathbf{H}\mathbf{W}^{-1/2}\mathbf{W}^{1/2}\mathbf{L}^{-1}\mathbf{W}^{1/2}\mathbf{W}^{-1/2}\mathbf{C}\mathbf{W}^{-1/2}\mathbf{W}^{1/2}\mathbf{L}^{-1}\mathbf{W}^{1/2} \times \cdots \\
&\quad \cdots \times \mathbf{W}^{-1/2}\mathbf{H}\mathbf{W}^{-1/2}\mathbf{W}^{1/2}\mathbf{A} \\
&= h^2 tr\mathbf{D}_\lambda\mathbf{P}'\mathbf{P}\mathbf{D}_{\frac{1}{1+\lambda h}}\mathbf{P}'\mathbf{W}^{-1/2}\mathbf{C}\mathbf{W}^{-1/2}\mathbf{P}\mathbf{D}_{\frac{1}{1+\lambda h}}\mathbf{P}'\mathbf{P}\mathbf{D}_\lambda\mathbf{P}'\mathbf{W}^{1/2}\mathbf{A}\mathbf{W}^{1/2}\mathbf{P} && (31) \\
&= h^2 tr\mathbf{D}_{\frac{\lambda}{1+\lambda h}}[\gamma_{ij}]\mathbf{D}_{\frac{\lambda}{1+\lambda h}}[\alpha_{ij}] && (32) \\
&= h^2 tr\left[\frac{\gamma_{ij}\lambda_i}{1+\lambda_i h}\right]\left[\frac{\alpha_{ij}\lambda_i}{1+\lambda_i h}\right] \\
&= h^2\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\gamma_{ij}\lambda_i}{1+\lambda_i h}\cdot\frac{\alpha_{ji}\lambda_j}{1+\lambda_j h} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\gamma_{ij}\alpha_{ij}\lambda_i\lambda_j h^2}{(1+\lambda_i h)(1+\lambda_j h)} && (33)
\end{aligned}$$

The symbol $\mathbf{D}_{\frac{\lambda}{1+\lambda h}}$ denotes a diagonal matrix whose ith diagonal element is $\frac{\lambda_i}{1+\lambda_i h}$.

Equation (31) follows from (24), (30) and (18). Equation (32) follows from (13), (11), and the fact that $\mathbf{P}$ is orthogonal. The remaining steps are analogous to those use to get (27), (28), and (29).

Adding (29) and (33) completes the proof.

Table 1
Data and Modified Graduations for Example 1

					Modified Whittaker			
age	*deaths*	*exposure*	*crude rate*	*theta*	$z = 1$	$z = 2$	$z = 3$	$z = 4$
15	190	372549020	0.67	0.51	0.77	0.85	0.73	0.67
16	505	515306122	0.81	0.98	0.93	1.01	0.92	0.88
17	749	571755725	0.95	1.31	1.08	1.16	1.11	1.09
18	670	614678899	1.07	1.09	1.21	1.30	1.27	1.27
19	1111	653529412	1.16	1.70	1.30	1.40	1.39	1.41
20	828	701694915	1.19	1.18	1.33	1.43	1.44	1.46
21	1013	803968254	1.16	1.26	1.29	1.39	1.42	1.44
22	1354	820606061	1.13	1.65	1.26	1.35	1.38	1.41
23	1071	811363636	1.11	1.32	1.23	1.31	1.35	1.38
24	1066	795522388	1.10	1.34	1.20	1.28	1.32	1.35
25	1121	789436620	1.08	1.42	1.16	1.24	1.28	1.30
26	1124	780555556	1.07	1.44	1.13	1.20	1.24	1.25
27	778	793877551	1.04	0.98	1.07	1.14	1.17	1.18
28	929	829464286	1.09	1.12	1.09	1.16	1.18	1.19
29	1137	895275591	1.13	1.27	1.10	1.16	1.18	1.19
30	1375	1082677165	1.18	1.27	1.12	1.17	1.19	1.19
31	1288	1238461538	1.23	1.04	1.13	1.17	1.19	1.19
32	1604	1485185185	1.25	1.08	1.11	1.15	1.16	1.15
33	2040	1906542056	1.28	1.07	1.10	1.13	1.14	1.13
34	2559	2347706422	1.33	1.09	1.12	1.14	1.13	1.12
35	3479	2851639344	1.40	1.22	1.15	1.15	1.15	1.14
36	4402	3285074627	1.50	1.34	1.20	1.20	1.18	1.18
37	4556	3615873016	1.63	1.26	1.27	1.26	1.25	1.24
38	5426	4019259259	1.79	1.35	1.35	1.35	1.34	1.33
39	6262	4409859155	1.98	1.42	1.46	1.46	1.45	1.45
40	7778	4861250000	2.20	1.60	1.60	1.60	1.59	1.59
41	9104	5293023256	2.43	1.72	1.74	1.74	1.74	1.75
42	10274	5583695652	2.68	1.84	1.90	1.91	1.91	1.92
43	11752	5817821782	2.96	2.02	2.09	2.10	2.11	2.12
44	13948	6064347826	3.27	2.30	2.32	2.33	2.33	2.34
45	16309	6296911197	3.62	2.59	2.58	2.59	2.59	2.60
46	18552	6625714286	3.99	2.80	2.86	2.87	2.87	2.87
47	23323	7003903904	4.42	3.33	3.20	3.19	3.19	3.19
48	26808	7385123967	4.92	3.63	3.58	3.57	3.56	3.56
49	31918	7728329298	5.51	4.13	4.01	4.01	4.01	4.01

Deaths are in units of $1000.
Rates have been multiplied by 1000.

(Continued on next page.)

Table 1 (Continued)
Data and Modified Graduations for Example 1

age	deaths	exposure	crude rate	theta	Modified Whittaker $z=1$	$z=2$	$z=3$	$z=4$
50	34926	7992219680	6.17	4.37	4.47	4.50	4.50	4.50
51	41151	8230200000	6.84	5.00	4.96	4.98	4.99	4.98
52	45272	8352767528	7.50	5.42	5.43	5.45	5.45	5.45
53	49473	8371065990	8.23	5.91	5.96	5.96	5.96	5.96
54	57868	8386666667	9.05	6.90	6.57	6.53	6.53	6.53
55	59556	8352875175	10.03	7.13	7.23	7.22	7.22	7.22
56	65696	8191521197	11.17	8.02	8.03	8.02	8.03	8.04
57	71306	7870419426	12.42	9.06	8.93	8.92	8.94	8.94
58	73390	7597308489	13.73	9.66	9.83	9.86	9.88	9.89
59	74447	7186003861	15.10	10.36	10.81	10.87	10.89	10.89
60	78935	6735068259	16.50	11.72	11.92	11.93	11.92	11.92
61	86003	6461532682	18.01	13.31	13.18	13.09	13.06	13.06
62	90755	6078700603	19.69	14.93	14.56	14.41	14.36	14.35
63	92295	5693707588	21.63	16.21	16.10	15.95	15.89	15.89
64	94658	5282254464	23.81	17.92	17.82	17.70	17.65	17.65
65	86192	4482163287	26.17	19.23	19.67	19.63	19.59	19.60
66	86877	4057776740	28.73	21.41	21.76	21.75	21.74	21.76
67	87659	3728583581	31.40	23.51	24.00	24.02	24.02	24.04
68	91204	3482397862	34.21	26.19	26.43	26.45	26.46	26.48
69	90921	3236774653	36.99	28.09	28.87	28.90	28.91	28.92
70	94129	2987273881	39.92	31.51	31.55	31.53	31.53	31.52
71	99713	2755263885	43.46	36.19	34.79	34.74	34.73	34.69
72	95384	2506806833	47.47	38.05	38.35	38.38	38.37	38.30
73	94812	2275852136	51.73	41.66	42.18	42.27	42.25	42.15
74	93133	2063660536	56.43	45.13	46.47	46.55	46.54	46.40
75	91263	1830018047	61.64	49.87	51.35	51.29	51.29	51.12
76	87075	1132314694	67.41	76.90	56.86	56.50	56.53	56.36
77	84758	1411222111	73.71	60.06	62.19	62.12	62.23	62.08
78	80831	1227688335	80.63	65.84	68.17	68.25	68.47	68.36
79	77098	1068727474	88.00	72.14	74.67	74.77	75.10	75.07
80	73066	943517562	95.60	77.44	81.47	81.49	81.91	81.98
81	73011	819427609	103.29	89.10	88.50	88.28	88.77	88.97
82	69436	700736704	111.63	99.09	96.13	95.66	96.20	96.55
83	58607	583676925	120.51	100.41	104.21	103.55	104.09	104.61
84	52998	476858017	130.63	111.14	113.56	112.59	113.10	113.80

Deaths are in units of \$1000.
(Continued on next page.)
Rates have been multiplied by 1000.

Table 1 (Continued)
Data and Modified Graduations for Example 1

					Modified Whittaker			
age	*deaths*	*exposure*	*crude rate*	*theta*	$z = 1$	$z = 2$	$z = 3$	$z = 4$
85	43008	373106619	141.76	115.27	123.91	122.59	123.00	123.88
86	40299	297827212	153.97	135.31	135.43	133.63	133.85	134.87
87	33557	231061076	167.26	145.23	148.02	145.68	145.61	146.74
88	27862	176419933	181.42	157.93	161.51	158.57	158.09	159.21
89	22441	134805070	193.63	166.47	173.19	169.59	168.56	169.52
90	20069	103887566	202.96	193.18	182.17	177.86	176.11	176.72
91	14489	73915927	210.32	196.02	189.24	184.20	181.59	181.58
92	11165	52834564	217.76	211.32	196.38	190.57	186.90	185.94
93	8632	39120780	230.07	220.65	208.20	201.52	196.54	194.22
94	6503	27181909	248.83	239.24	226.26	218.58	211.97	207.73
95	5432	18098824	267.93	300.13	244.70	235.95	227.42	220.59
96	2509	7440247	279.78	337.22	256.11	246.34	235.67	225.59
97	783	3804665	292.83	205.80	268.68	257.85	244.70	230.55
98	633	2018302	306.15	313.63	281.55	269.59	253.64	234.47
99	940	13599537	319.74	69.12	294.70	281.60	262.48	237.23
100	123	480000	333.56	256.25	308.22	293.83	271.17	238.71

Deaths are in units of \$1000.
Rates have been multiplied by 1000.

Table 2
Optimal Values of h and *Bayes Risk*

Example 1			Example 2		
z	h	*Bayes Risk*	z	h	*Bayes Risk*
1	10.327	0.00020895	1	7.552	0.00408858
2	103.381	0.00023696	2	37.265	0.00490776
3	1226.896	0.00026329	3	303.221	0.00546794
4	16081.602	0.00028290	4	2725.891	0.00584935

Table 3
Data and Modified Graduations for Example 2

					Modified Whittaker			
age	*deaths*	*exposure*	*crude rate*	*theta*	$z=1$	$z=2$	$z=3$	$z=4$
20	1	111	1.48	9.01	0.73	1.45	2.92	4.71
21	2	532	1.50	3.76	0.67	0.99	1.51	1.93
22	0	476	1.53	0.00	0.48	0.64	0.72	0.68
23	0	324	1.49	0.00	0.35	0.36	0.29	0.16
24	0	279	1.44	0.00	0.26	0.20	0.09	0.02
25	0	289	1.38	0.00	0.21	0.11	0.02	0.00
26	0	285	1.32	0.00	0.18	0.07	0.01	0.00
27	0	261	1.21	0.00	0.15	0.04	0.00	0.00
28	0	230	1.13	0.00	0.15	0.04	0.00	0.00
29	0	210	1.09	0.00	0.17	0.06	0.02	0.01
30	0	256	1.11	0.00	0.24	0.12	0.07	0.06
31	0	338	1.13	0.00	0.33	0.21	0.15	0.14
32	2	558	1.15	3.58	0.49	0.32	0.25	0.25
33	1	1193	1.16	0.84	0.50	0.41	0.36	0.37
34	0	1452	1.19	0.00	0.47	0.50	0.48	0.49
35	3	1773	1.23	1.69	0.69	0.61	0.60	0.62
36	1	2127	1.28	0.47	0.71	0.71	0.72	0.75
37	3	2745	1.36	1.09	0.85	0.83	0.87	0.91
38	2	2767	1.48	0.72	0.92	0.99	1.06	1.11
39	2	2464	1.62	0.81	1.11	1.21	1.29	1.34
40	4	2370	1.78	1.69	1.47	1.50	1.56	1.60
41	4	2312	1.95	1.73	1.80	1.83	1.87	1.87
42	7	2310	2.14	3.03	2.21	2.18	2.19	2.16
43	5	2062	2.36	2.42	2.36	2.52	2.52	2.45
44	2	1918	2.61	1.04	2.51	2.88	2.83	2.72
45	8	1935	2.89	4.13	3.32	3.26	3.10	2.96
46	13	1753	3.19	7.42	3.96	3.54	3.29	3.15
47	8	1584	3.53	5.05	3.79	3.61	3.43	3.33
48	2	1581	3.93	1.27	3.37	3.59	3.56	3.54
49	7	1471	4.40	4.76	3.71	3.68	3.75	3.81
50	4	1518	4.93	2.64	3.84	3.89	4.04	4.18
51	7	1375	5.47	5.09	4.32	4.24	4.45	4.63
52	4	1345	6.02	2.97	4.62	4.77	5.01	5.19
53	4	1306	6.58	3.06	5.35	5.56	5.74	5.88
54	11	1238	7.24	8.89	6.85	6.69	6.72	6.78
55	11	1210	8.02	9.09	8.17	8.02	7.91	7.87
56	13	1120	9.03	11.61	9.61	9.55	9.35	9.22

Rates have been multiplied by 1000. (Continued on next page.)

Table 3 (Continued)
Data and Modified Graduations for Example 2

age	*deaths*	*exposure*	*crude rate*	*theta*	Modified Whittaker $z=1$	$z=2$	$z=3$	$z=4$
57	12	1054	10.07	11.39	10.75	10.97	10.75	10.55
58	12	1161	11.12	10.34	11.79	12.19	11.99	11.78
59	19	1028	12.18	18.48	13.14	13.19	13.00	12.84
60	12	951	13.25	12.62	13.65	13.85	13.76	13.70
61	16	861	14.40	18.58	14.37	14.34	14.36	14.43
62	12	756	15.75	15.87	14.67	14.81	14.98	15.17
63	6	696	17.30	8.62	14.96	15.42	15.71	15.98
64	10	599	19.10	16.69	16.28	16.42	16.68	16.97
65	6	430	21.38	13.95	17.97	18.07	18.19	18.40
66	5	415	24.01	12.05	20.28	20.31	20.16	20.22
67	8	332	26.93	24.10	23.55	23.09	22.59	22.42
68	13	326	30.05	39.88	27.07	26.20	25.39	24.98
69	11	266	33.28	41.35	30.09	29.39	28.47	27.85
70	7	270	36.57	25.93	32.68	32.55	31.77	31.01
71	12	241	39.88	49.79	35.59	35.70	35.23	34.42
72	4	223	43.23	17.94	37.97	38.85	38.84	38.10
73	7	173	46.64	40.46	41.29	42.11	42.61	42.03
74	6	136	50.78	44.12	45.35	46.05	47.11	46.79
75	13	137	55.47	94.89	49.97	50.38	52.09	52.13
76	4	112	60.66	35.71	54.18	54.95	57.39	57.87
77	6	116	66.33	51.72	59.20	59.75	62.86	63.84
78	4	96	72.56	41.67	64.89	64.84	68.41	69.92
79	10	104	79.21	96.15	71.42	70.01	73.72	75.72
80	4	71	86.04	56.34	77.73	74.85	78.34	80.75
81	8	52	92.96	153.85	84.41	79.21	81.98	84.65
82	7	53	100.46	132.08	91.11	83.39	84.92	87.66
83	3	32	108.45	93.75	97.92	87.28	86.93	89.44
84	5	31	117.56	161.29	105.79	91.50	88.46	90.36
85	0	9	127.58	0.00	114.21	95.90	89.25	90.05
86	1	10	138.57	100.00	123.86	100.67	89.30	88.40
87	0	13	150.53	0.00	134.44	105.91	88.59	85.34
88	0	1	163.27	0.00	146.33	111.55	86.98	80.68
89	0	5	174.26	0.00	156.61	115.49	82.66	72.74
90	1	3	182.66	333.33	164.68	117.12	75.37	61.55
91	0	1	189.28	0.00	170.97	117.17	66.13	48.49
92	1	4	195.98	250.00	177.40	117.20	56.50	35.42
93	0	1	207.06	0.00	188.02	120.68	48.99	24.71

Rates have been multiplied by 1000.

FIGURE 1
MSE of v(0.8) and v(1.0)

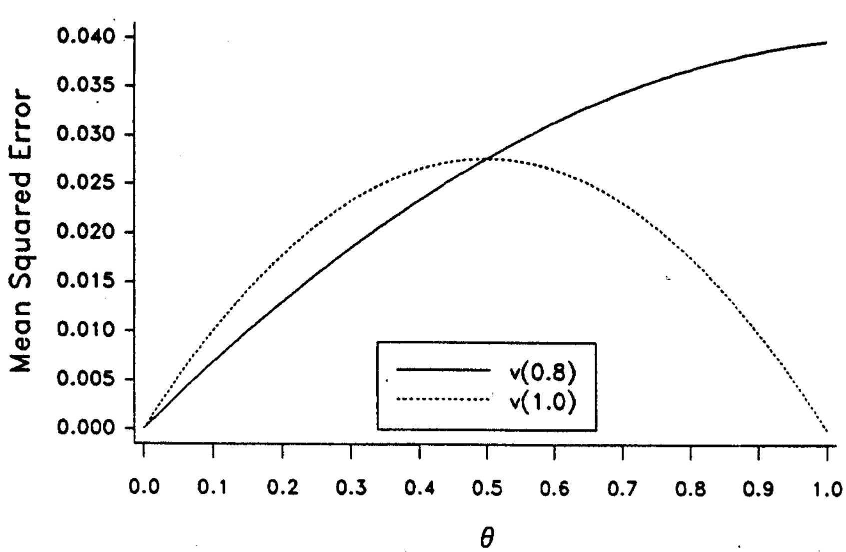


FIGURE 2a
Standard Whittaker Graduations
Example 1, h = infinity

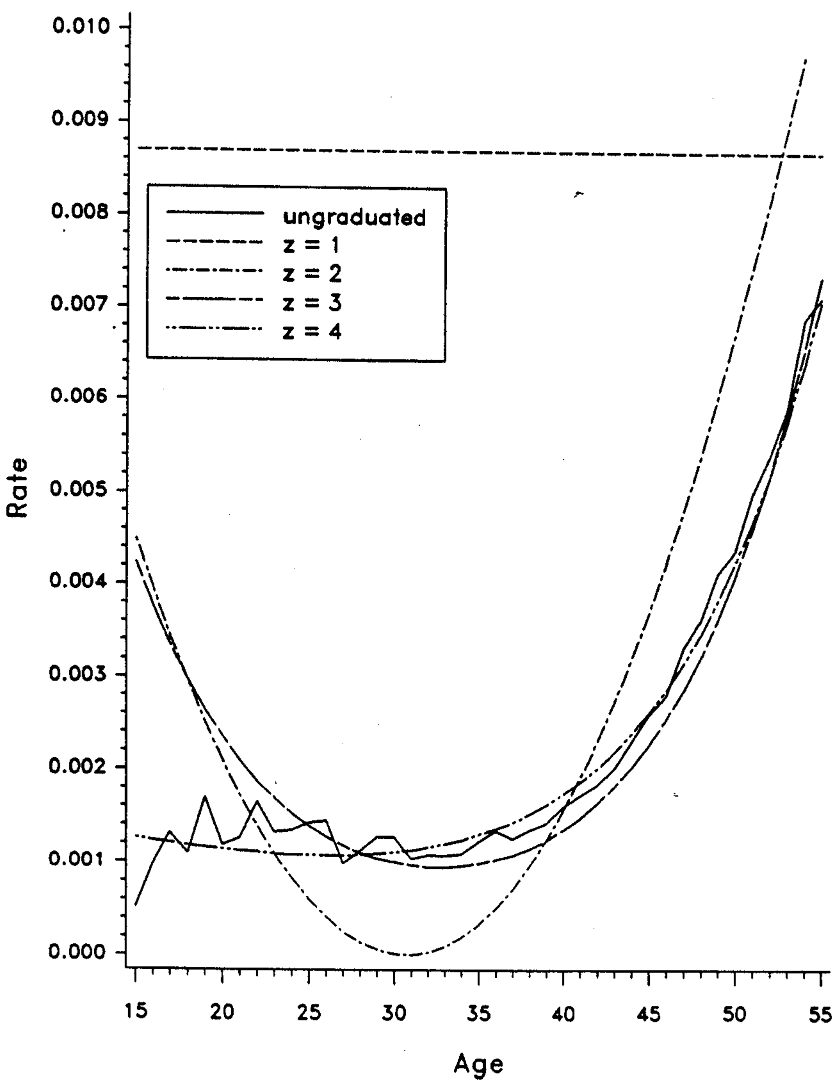
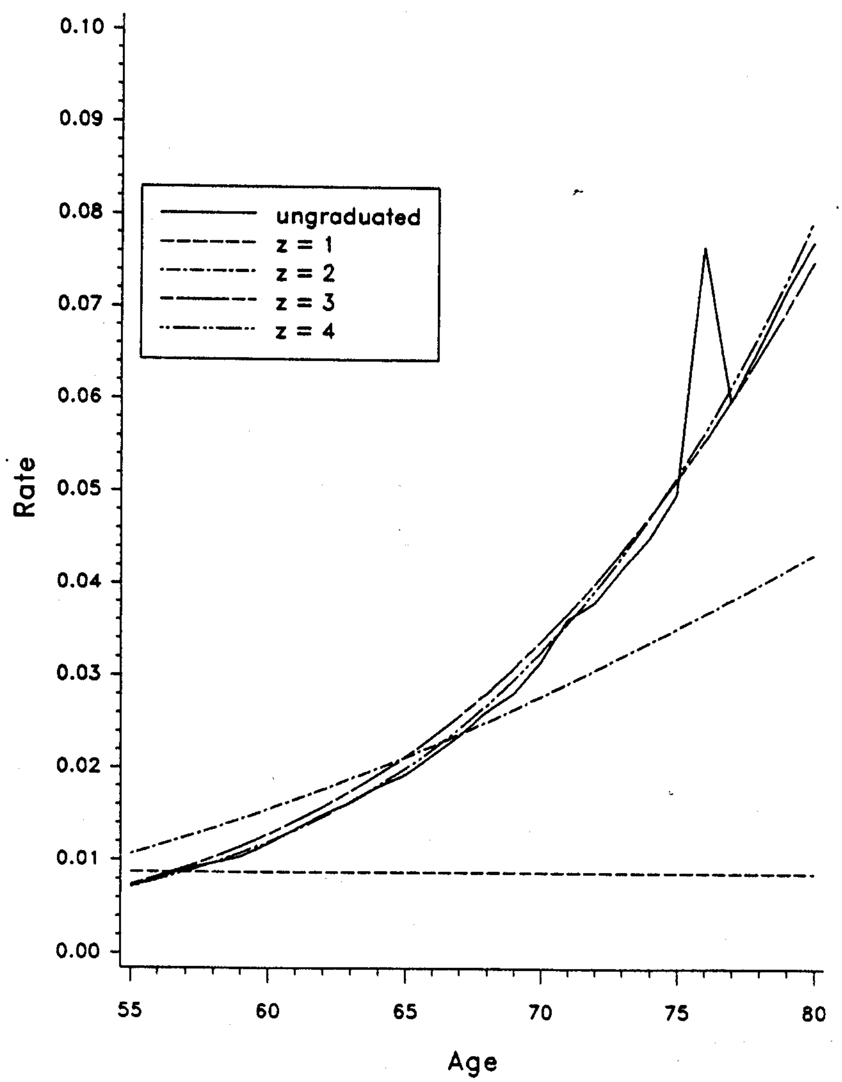


FIGURE 2b
Standard Whittaker Graduations
Example 1, h = infinity



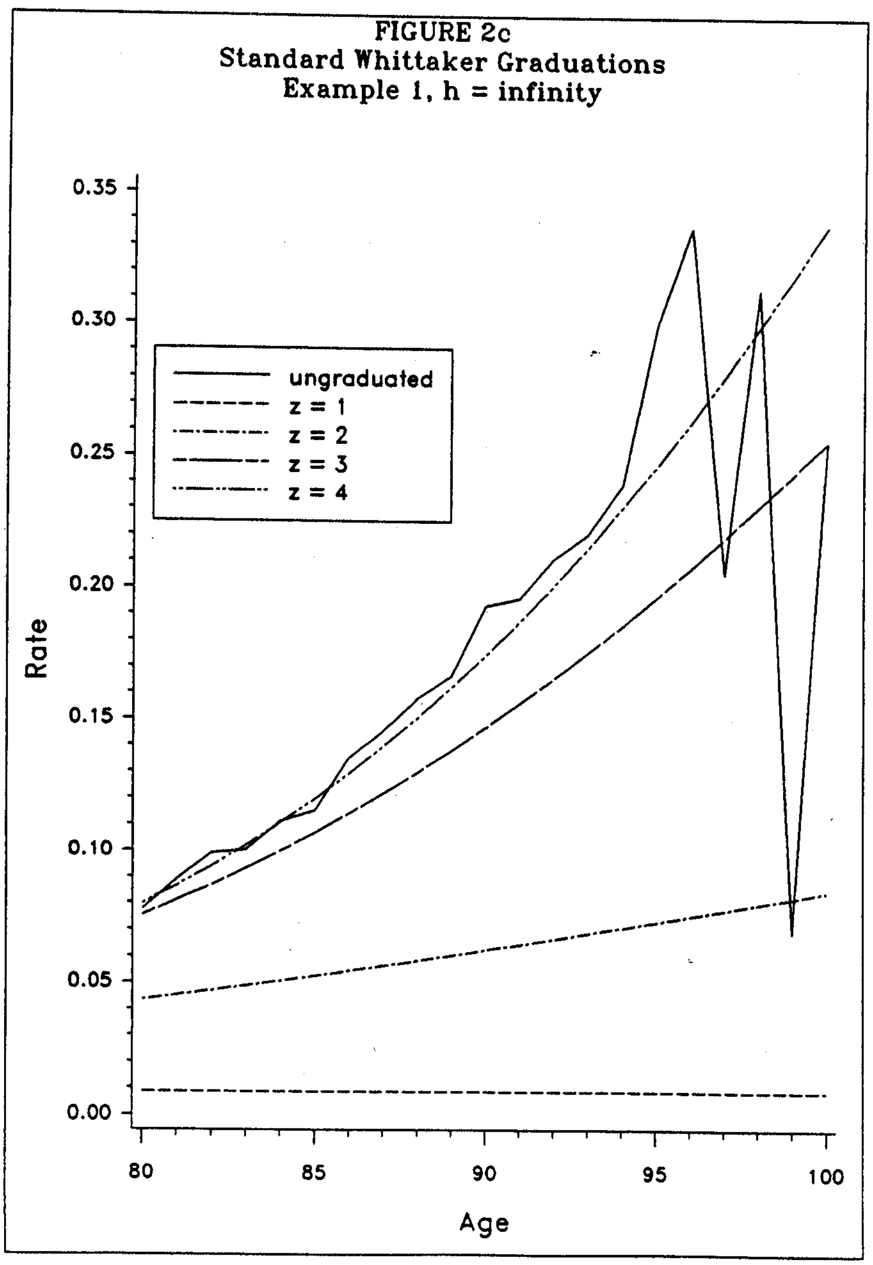


FIGURE 2c
Standard Whittaker Graduations
Example 1, h = infinity
ungraduated
z = 1
z = 2
z = 3
z = 4
Rate
Age
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
80
85
90
95
100

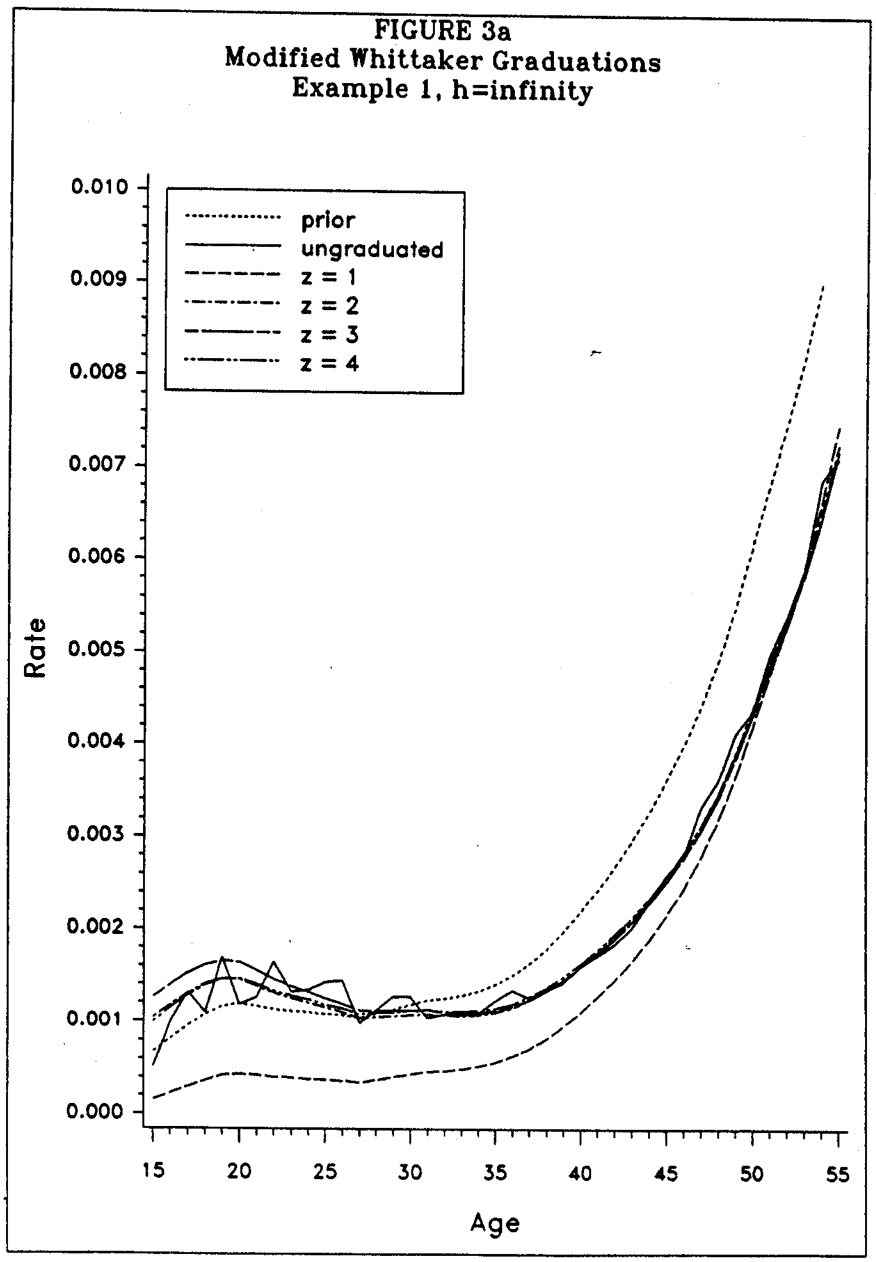


FIGURE 3a
Modified Whittaker Graduations
Example 1, h=infinity
prior
ungraduated
z = 1
z = 2
z = 3
z = 4
Rate
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
15
20
25
30
35
40
45
50
55
Age

FIGURE 3b
Modified Whittaker Graduations
Example 1, h=infinity

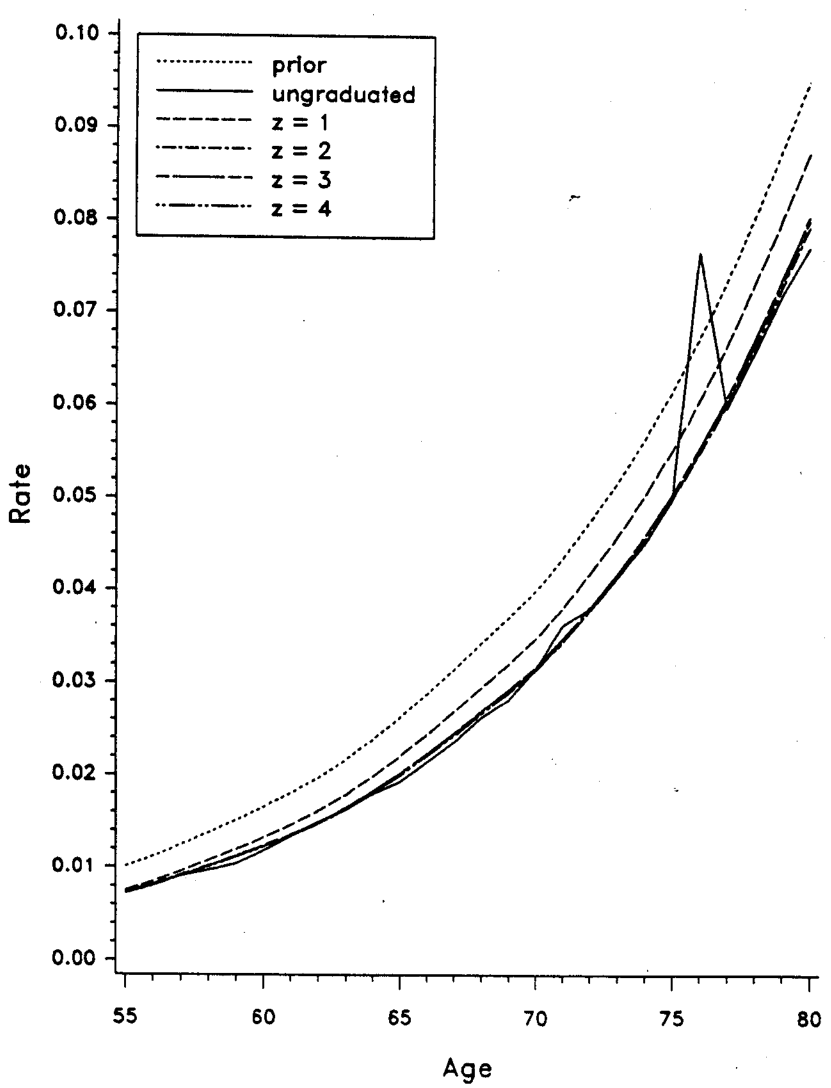


FIGURE 3c
Modified Whittaker Graduations
Example 1, h=infinity

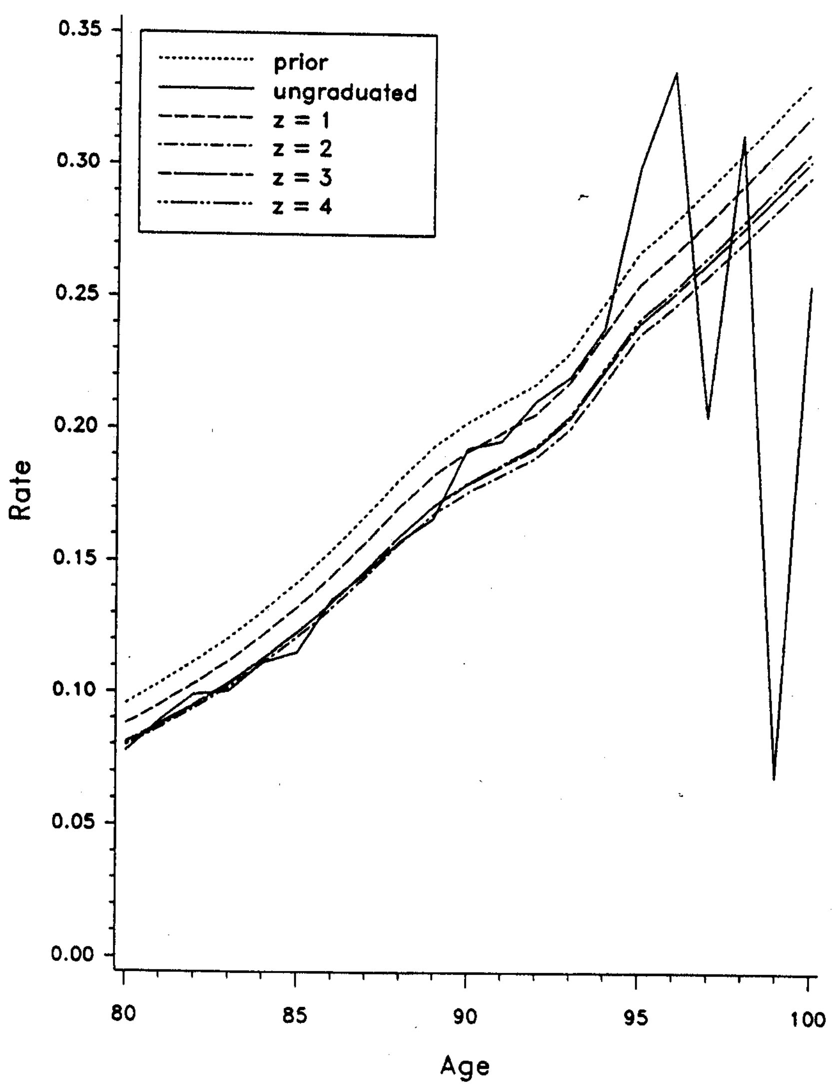
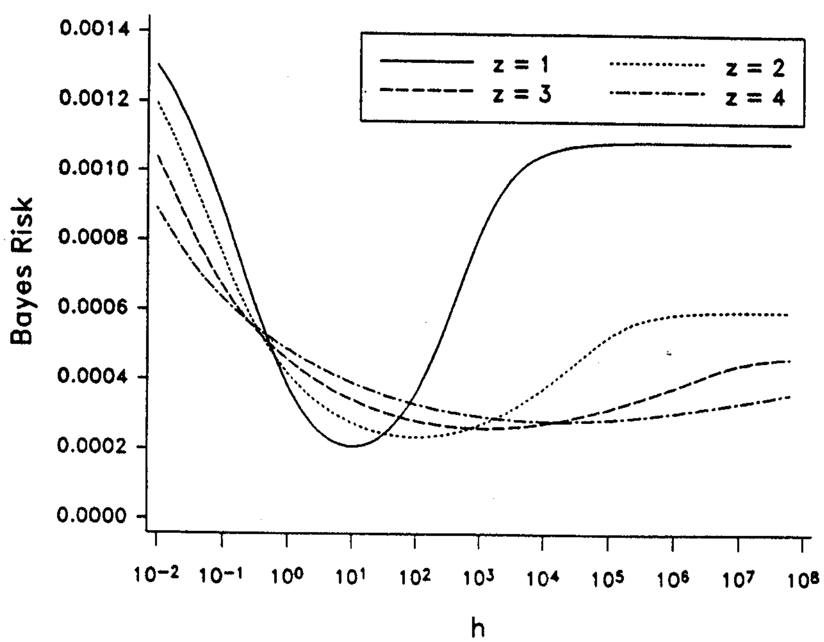


FIGURE 4
Bayes Risk vs. h



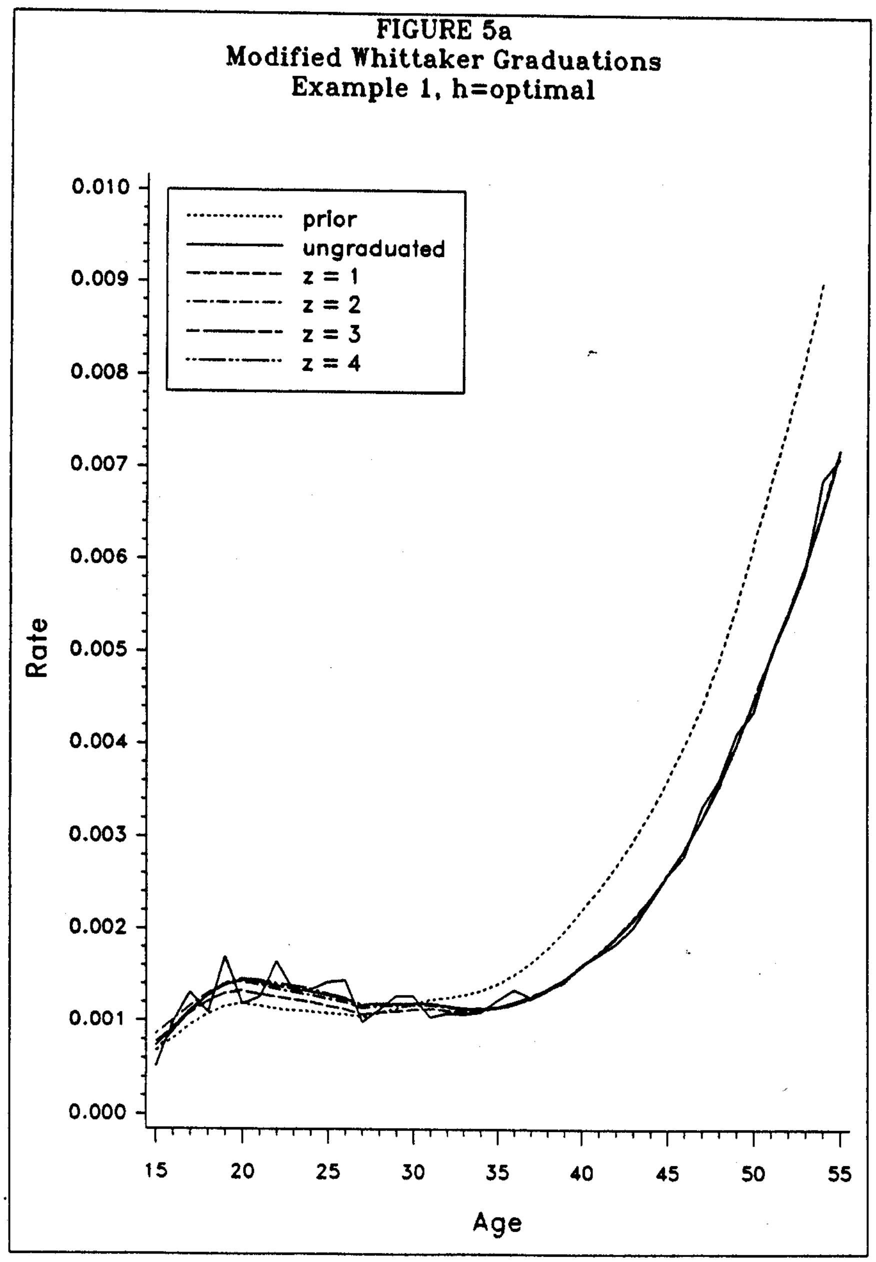


FIGURE 5a
Modified Whittaker Graduations
Example 1, h=optimal
prior
ungraduated
z = 1
z = 2
z = 3
z = 4
Rate
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
15
20
25
30
35
40
45
50
55
Age

FIGURE 5b
Modified Whittaker Graduations
Example 1, h=optimal

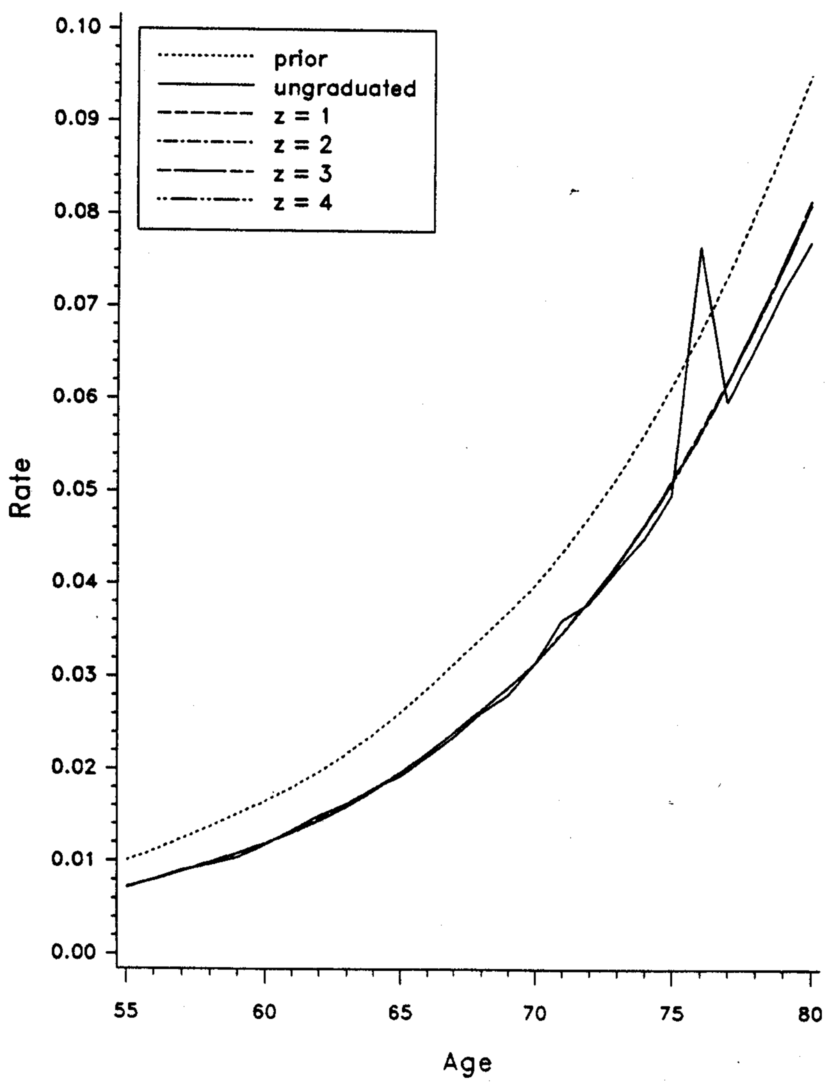
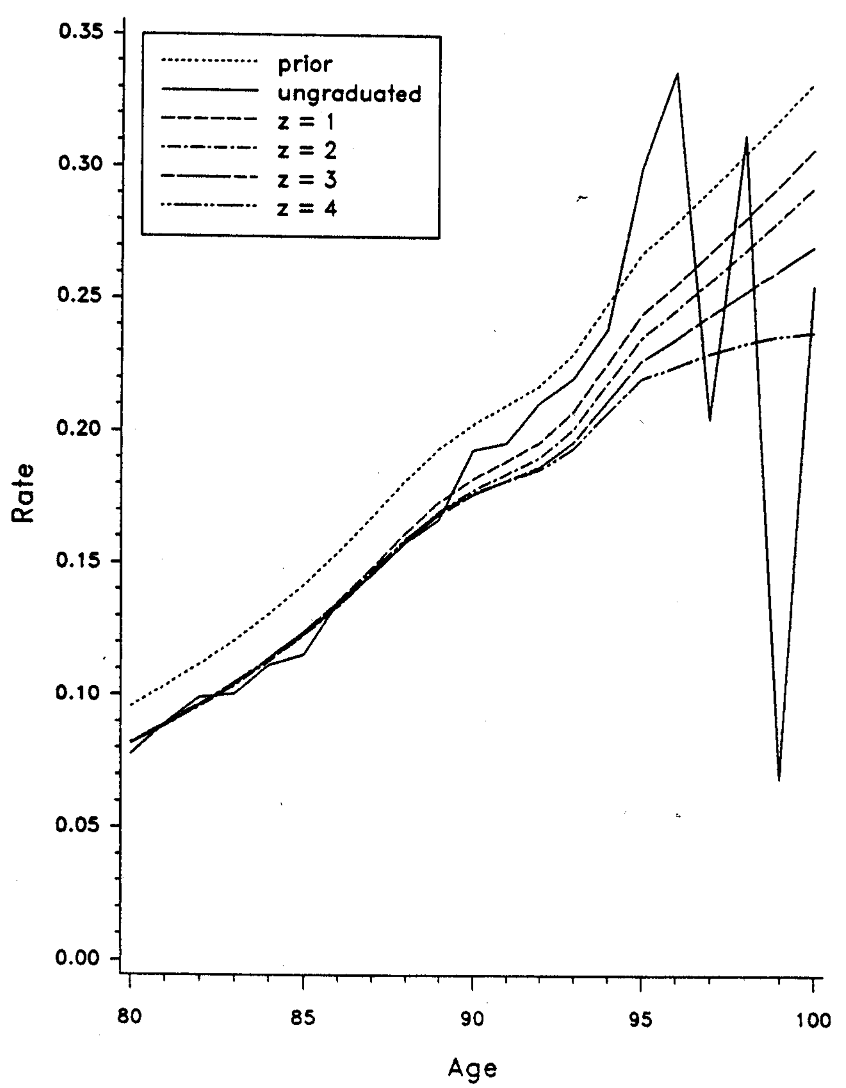


FIGURE 5c
Modified Whittaker Graduations
Example 1, h=optimal



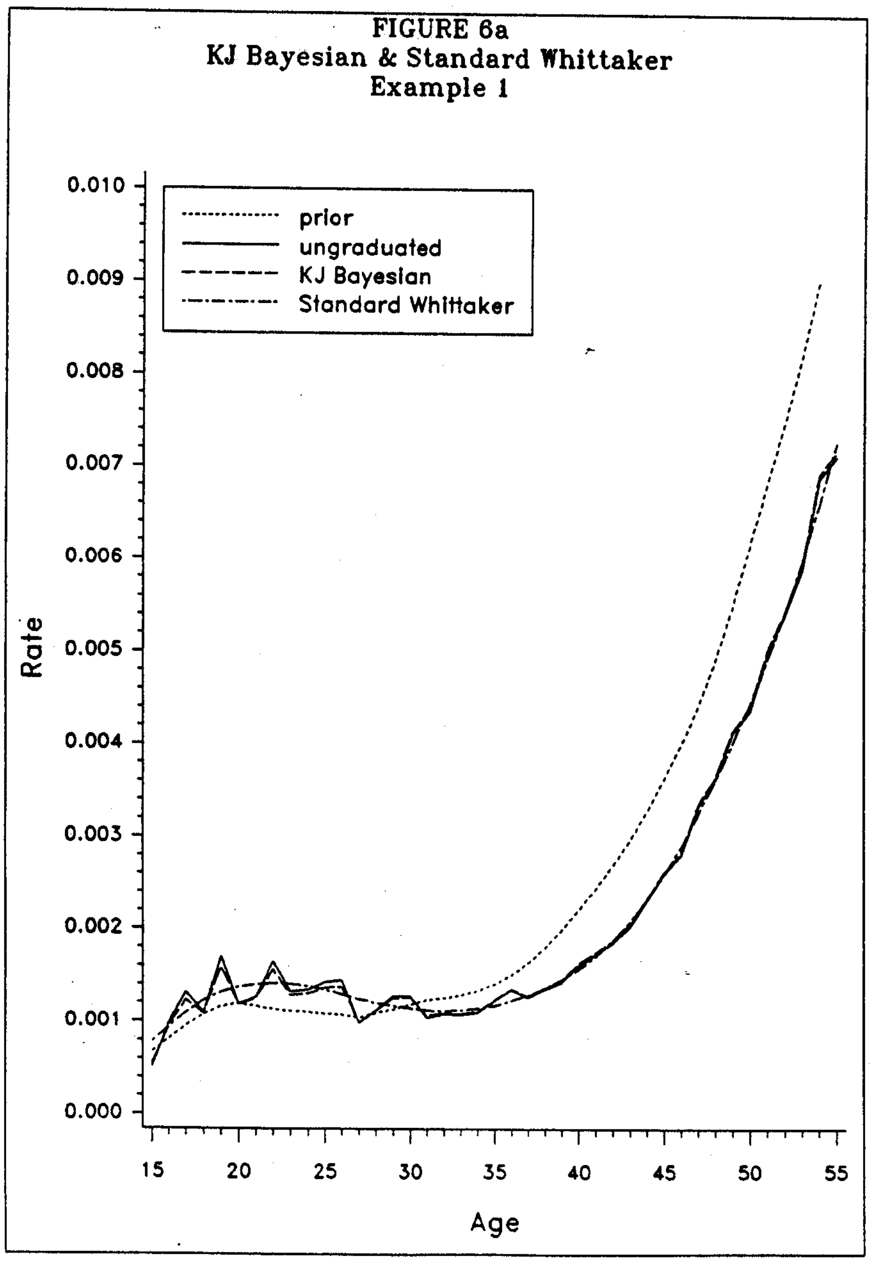


FIGURE 6a
KJ Bayesian & Standard Whittaker
Example 1
prior
ungraduated
KJ Bayesian
Standard Whittaker
Rate
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
15
20
25
30
35
40
45
50
55
Age

FIGURE 6b
KJ Bayesian & Standard Whittaker
Example 1

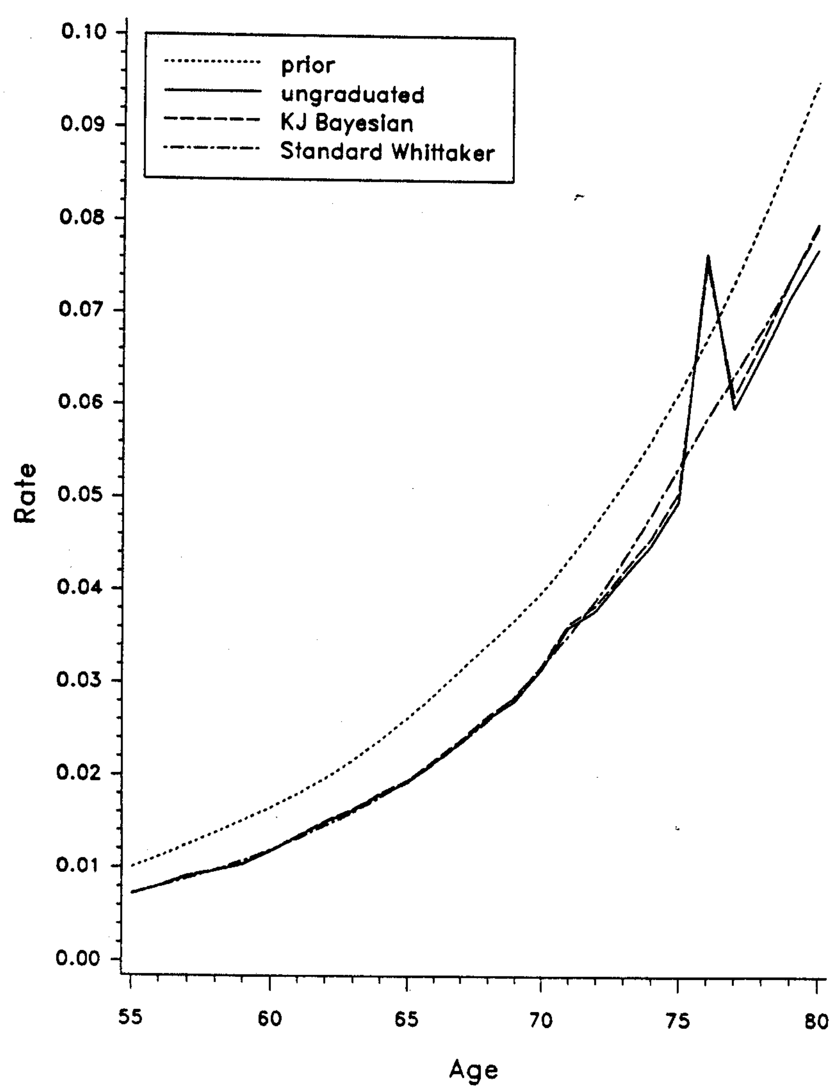


FIGURE 6c
KJ Bayesian & Standard Whittaker
Example 1

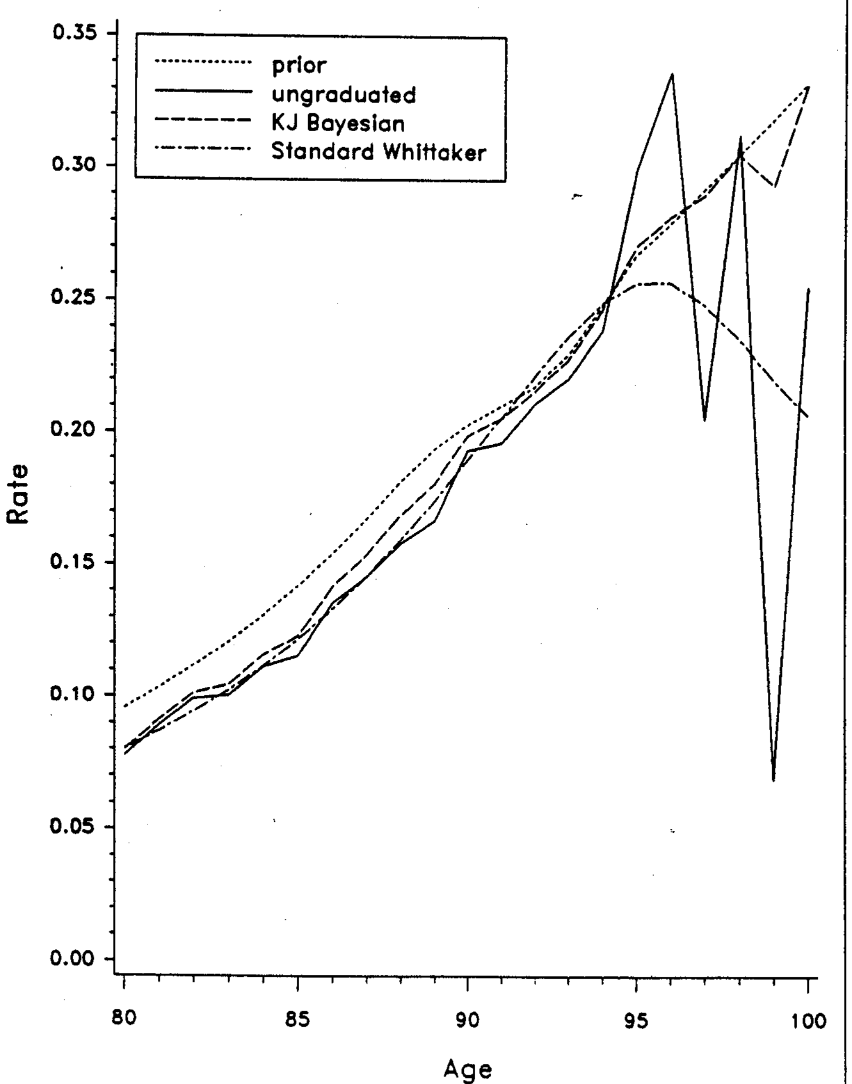


FIGURE 7a
Modified Whittaker Graduations
Example 2, h=optimal

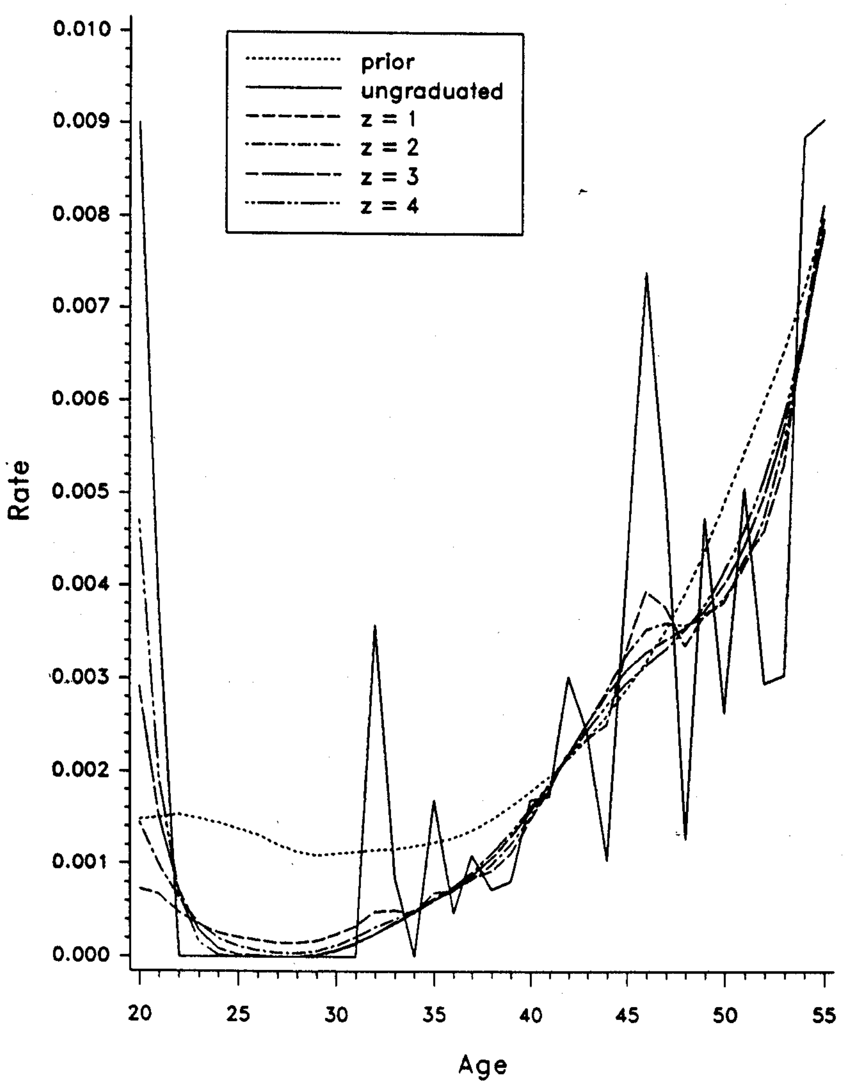


FIGURE 7b
Modified Whittaker Graduations
Example 2, h=optimal

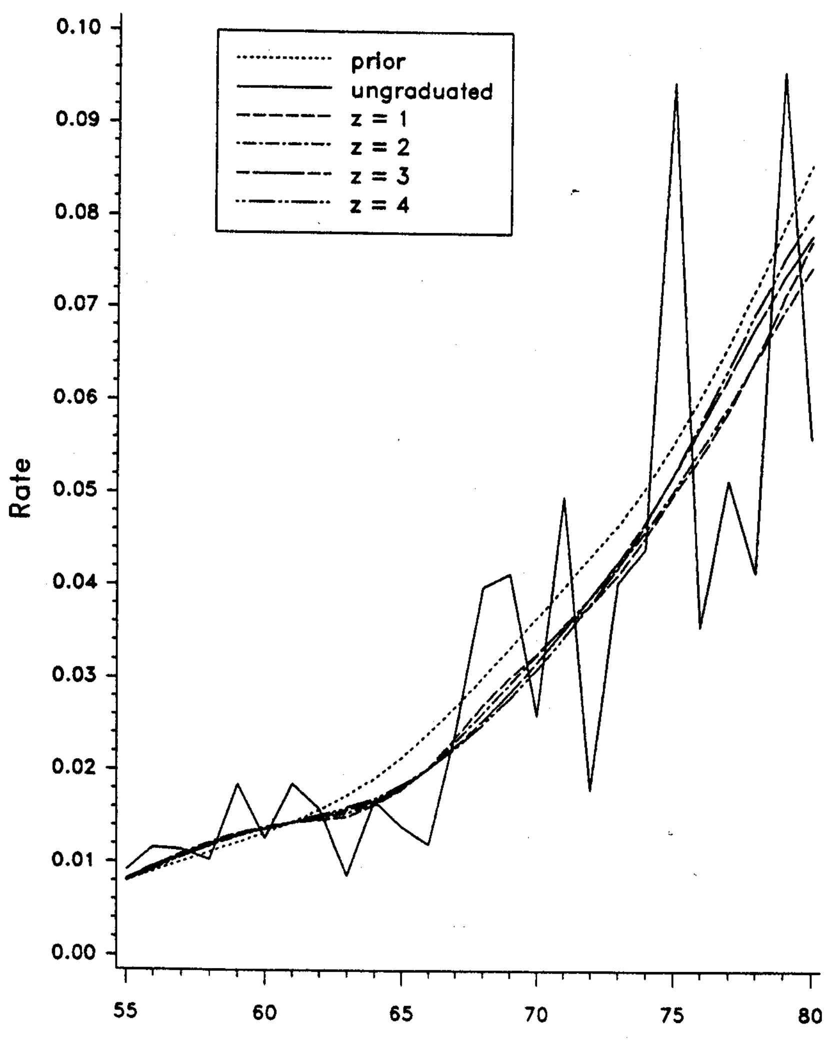


FIGURE 7c
Modified Whittaker Graduations
Example 2, h=optimal