Bayesian Bivariate Graduation and Forecasting

by

James C. Hickman

Robert B. Miller

ABSTRACT

The estimation of surivival functions is fundamental to the disciplines of reliability engineering, biostatistics, demography, and actuarial science. In actuarial applications we deal with populations of insureds, annuitants, and pensioners. We need to estimate probabilities of individuals remaining in the populations and moving from the populations for reasons of death, change in health status, voluntary withdrawal, etc. Estimates of these probabilities aid us in premium and reserve determination and, as a consequence, in developing investment strategies and cash flow projections.

Let there be K age groups in a life table. Suppose that for each age group a death rate has been observed for each of c_1 calendar periods. We present a Bayesian approach to (1) estimation of the underlying death rates for the observation period (graduation), (2) estimation of the underlying death rates for c_2 future calendar periods (extrapolation), and (3) prediction of the observed death rates for the c_2 future calendar periods (forecasting).

KEY WORDS: Bayesian, Graduation, Forecasting, Mortality.

1. INTRODUCTION

The estimation of survival functions is fundamental to the disciplines of reliability engineering, biostatistics, demography, and actuarial science. In actuarial applications we deal with populations of insureds (lives or property), annuitants, or pensioners. We need to estimate probabilities of items remaining in populations and moving from populations for reasons of death or damage, change in health status, voluntary withdrawal, or what have you. Our estimates of these probabilities, along with other information, aid us in premium and reserve determination and, as a consequence in developing investment strategies and cash flow projections. In demography estimates of mortality, fertility, and marriage rates are used to obtain predictions of age distributions and sizes of populations, which are needed for a variety of planning and policy purposes.

In this paper we shall present our results in terms of the estimation of human mortality rates, but the results carry over to other applications. We shall discuss a Bayesian model that encompasses both graduation (smoothing) and prediction.

2. LITERATURE REVIEW

Estimates of mortality rates are almost universally graduated or adjusted to conform more nearly to a priori smoothness characteristics (Hoem (1972), (1976) and Miller (1942)). The deviation of the estimates from expected behavior is the result of several sources of error. The data may be incomplete either because they are a sample or because individuals become unobservable during the observation period (censoring). The estimation methods themselves usually rely on approximations that introduce some degree of error. Finally, reporting and processing errors can be quite serious.

Adjustment of raw estimates can be done by fitting them to smooth functions such as Gompertz, Makeham, or Hadwiger (inverse Gaussian) functions (Hoem (1972), (1976)). A more common practice in actuarial science is to find adjusted estimates that minimize an objective function containing "fit" and "smoothness" components, i.e.,

$$\min_{\underline{v}} \left\{ \sum_{x=1}^{K} w_x (u_x - v_x)^{q_1} + \theta \sum_{x=1}^{K-m} (\Delta^m v_x)^{q_2} \right\},$$

where

K = number of age groups (assumed "equally spaced")

x = age group index

u_x = raw estimate of mortality rate x

v_x = graduated estimate of mortality rate x

w_x = weight (usually the exposure to risk of death)

Δ^m = m-th forward difference operator

q_1, q_2 = positive numbers, usually integers and usually equal

θ = parameter measuring relative emphasis on "smoothness" over "fit"

E.T. Whittaker (1923) developed the latter formulation of the graduation problem using a Bayesian argument in which the likelihood function was proportional to

$$\exp\{ -\sum_{x=1}^{K} w_x(u_x-v_x)^2\},$$

and the prior density was proportional to

$$\exp\{ -\theta\sum_{x=1}^{K-2} (\Delta^2 v_x)^2\}.$$

He chose the mode of the resulting posterior distribution as the graduated values on the theory that the purpose of graduation was to obtain "most probable" death rates.

This prior density was singular on $\underline{u} = (v_1,\ldots,v_K)'$, and while singularity is not a disastrous quality for a prior, the justification for its use in this case is not immediately apparent. Whittaker may not have thought in terms of a prior on $\underline{v}$ at all but rather in terms of a prior on S of the form $\theta e^{-\theta S}$, where S could be any of a number of measures of smoothness. Many other possible prior functions come to mind, but Whittaker's choice has become more or less enshrined in actuarial practice.

I.J. Schoenberg (1964) modified Whittaker's objective function by allowing unequally-spaced arguments $a = x_1 \leq \ldots \leq x_n = b$, choosing $w_x = 1$, and treating v_x as a function having a square integrable derivative of order m. He showed that the solution of the problem

$$\min_{v}\{\sum_{\nu=1}^{n}(u_{x_\nu}-v_{x_\nu})^2 + \frac{\theta}{(m!)^2}\int_a^b (v_x^{(m)})^2 dx\}$$

was a unique spline function of order 2m with knots $x_\nu (1\leq\nu\leq n)$. He also showed that for the optimum spline function v_x, $(m!)^{-2}\int_a^b (v_x^{(m)})^2 dx$ could be written as a positive definite quadratic form in the m-th divided differences of v_x, the values depending on the knots. Wahba (1978) showed how a spline smoothing function could be obtained from a Bayesian argument using an improper prior. Marquardt (1974) rederived Whittaker's objective function using ridge regression arguments, which can also be given a Bayesian interpretation.

Kimeldorf and Jones (1967), using a direct Bayesian argument, proposed using nonsingular priors in the graduation problem and discussed the elicitation of prior parameters in some depth. (See also Dickey (1969).) Hickman and Miller (1978) suggested that prior specifications could be simplified by making variance stabilizing transformations on the raw estimates.

All the methods discussed so far were designed to smooth mortality rates arising from a single calendar period of observation. Our purpose in this paper is to develop a model that incorporates several calendar periods and that will be useful for prediction. McKay and Wilkin (1977) discussed a two-directional smoothing technique based on a direct extension of Whittaker's objective function, but their purpose was not prediction. Three papers on predicting demographic functions and reasons for doing so are Keyfitz (1972), Brass (1974), and Cox and Scott (1977).

In an extensive paper on mortality graduation and forecasting Cramér and Wold (1935) stated that the earliest attempt at forecasting mortality known to them was made by the Swedish astronomer Gyldén in 1878. The earliest cited commercial use of a mortality projection was 1901 and had to do with predicting annuity values for a pension fund. Cramér and Wold reported Swedish demographic data for males and females (exposed to risk and deaths) for the twelve five-year age intervals between the ages of 30 and 90 and for the 26 five-year calendar periods between 1800 and 1930. They predicted mortality rates through 1980.

3. MODEL SPECIFICATION

Definitions

Hoem (1971), (1972), and (1976), has presented a general stochastic model for demographic populations and discussed the estimation of transition rates (usually called forces in demographic and actuarial work). While we do not wish to reproduce Hoem's model, we need to present some notation and his key estimation theorem.

For any life in the population let the positive-valued random variable T stand for the time until the life dies. Assume that T has absolutely continuous distribution function $F(\cdot)$ with density $f(\cdot)$, and define the force of mortality (instantaneous death rate) to be

$$\mu(t) = f(t)/[1-F(t)].$$

for $t > 0$. If ω denotes the maximum value of T, then $\mu(\omega) = \infty$. If $F(t|x)$ denotes the conditional distribution function of time until death, given survival to time (age) x, then we have

$$F(t|x) = 1 - \exp\{-\int_0^t \mu(x+s)ds\}.$$

Similarly, we may define the force of transition from the population for reasons other than death, $\nu(t)$, say. Then the total force of transition is the sum $\mu(t) + \nu(t)$.

We break the age interval $[0,\omega)$ into subintervals $[0,x_1)$, $[x_1,x_2),\ldots,[x_{K+1},\omega)$ These intervals are typically 1, 3, 5, or 10 years in length. During an observation period, which would ordinarily be between 1 and 5 years in length, all the lives between ages x_k and x_{k+1} $(k = 0,1,\ldots,K)$ contribute "exposure to risk of death or withdrawal" during that age interval. The exposure in the age interval $[x_k,x_{k+1})$ is denoted by L_k and is the total time measured in years lived by individuals under study during the observation period. The exposures L_k and the numbers of deaths D_k during each age interval k are the data from which mortality estimates are made. In actuarial and demographic work we make estimates of $\mu(x_k^*)$, where x_k^* is a point in $[x_k,x_{k+1})$, graduate them, and then find other values of μ by interpolation. These values can then be used to estimate other important mortality functions such as ${}_tq_x = F(t|x)$.

Let n_k denote the total number of (stochastically independent) lives under study during a given observation period. Assume $L_k/n_k \xrightarrow{P} \tau_k > 0$ as $n_k \to \infty$, and take $\hat{\mu}(x_k^*) = D_k/L_k$. Let n denote the total number of lives ever under study. Then Hoem (1972) shows that, under suitable conditions, the quantities

$$\sqrt{n}[\hat{\mu}(x_k^*) - \mu(x_k^*)], \quad k = 1,\ldots,K,$$

are asymptotically independent and normally distributed with means 0 and asymptotic variances $\sigma_k^2 = \mu(x_k^*)/\tau_k$. It follows from a theorem in Rao [(1952), sec. 12e] that the quantities $\sqrt{\hat{\mu}(x_k^*)}$ are asymptotically independent and normally distributed with means $\sqrt{\mu(x_k^*)}$ and variances $1/4L_k$. These variances depend only on the observed quantity L_k, a very helpful simplification.

The Likelihood Function

Hoem's theorem continues to hold, under suitable conditions, when observations are made over c adjacent calendar periods. We shall index calendar periods by $\ell = 1,2,\ldots, c$, and we define (in fairly obvious notation) the quantities

$$u_{k\ell} = \sqrt{\hat{\mu}_\ell(x_{k\ell}^*) \times 1000}$$

$$v_{k\ell} = \sqrt{\mu_\ell(x_{k\ell}^*) \times 1000}$$

The force of mortality in the ℓ-th time period is $\mu_\ell(\cdot)$, and $x_{k\ell}^*$ is the estimation point in age interval $[x_k, x_{k+1})$ chosen during time period ℓ.

Let $\underline{u} = (u_{11}, u_{21}, \ldots, u_{K1}, u_{12}, \ldots, u_{Kc})'$, $\underline{v} = (v_{11}, v_{21}, \ldots, v_{K1}, v_{12}, \ldots, v_{Kc})'$, and $\underline{L}^{(-1)} = (L_{11}^{-1}, L_{21}^{-1}, \ldots, L_{K1}^{-1}, L_{12}^{-1}, \ldots, L_{Kc}^{-1})'$. Asymptotically, the conditional density of $\underline{u}$, given $\underline{v}$ and $\underline{L}$ is proportional to

$$\exp\{-\tfrac{1}{2}(\underline{u}-\underline{v})'B^{-1}(\underline{u}-\underline{v})\}, \qquad (1)$$

where $B = 250 \text{ diag } \underline{L}^{(-1)}$.

We let $c = c_1 + c_2$, where c_1 is the number of periods of observation and c_2 is the number of periods to be predicted. We partition $\underline{u}$, $\underline{v}$, and B as follows:

$$\underline{u} = \begin{bmatrix} \underset{Kc_1 \times 1}{\underline{u}^{(1)}} \\ \cdots \\ \underset{Kc_2 \times 1}{\underline{u}^{(2)}} \end{bmatrix}, \qquad \underline{v} = \begin{bmatrix} \underset{Kc_1 \times 1}{\underline{v}^{(1)}} \\ \cdots \\ \underset{Kc_2 \times 1}{\underline{v}^{(2)}} \end{bmatrix},$$

$$B = \begin{bmatrix} \underset{Kc_1 \times Kc_1}{B_{11}} & \vdots & \underset{Kc_1 \times Kc_2}{0} \\ \cdots & \vdots & \cdots \\ \underset{Kc_2 \times Kc_1}{0} & \vdots & \underset{Kc_2 \times Kc_2}{B_{22}} \end{bmatrix}$$

From a Bayesian point of view, given a prior distribution on $\underline{v}$ and data $(\underline{u}^{(1)}, \underline{L}^{(-1)})$, we wish to compute the posterior distributions of $\underline{v}^{(1)}$ and $\underline{v}^{(2)}$

and the predictive distribution of $\underline{u}^{(2)}$. The posterior distributions of $\underline{v}^{(1)}$ and $\underline{v}^{(2)}$ depend only on the elements of $\underline{L}^{(-1)}$ that come from the observation period, so the conditioning of the inferences appears natural. The predictive distribution of $\underline{u}^{(2)}$, given $(\underline{u}^{(1)},\underline{L}^{(-1)})$ requires a guess at the exposures that will be observed in the prediction period. A formal mechanism for guessing at these exposures is not part of our model, and we assume that the future exposures will be projected by standard demographic techniques. We shall base the necessary conditional distributions (likelihood functions) on the normal density in (1). The approximation error in doing so is likely to be negligible in most actuarial and demographic applications as the exposures tend to be large.

The Prior Distribution

The function of the prior in Bayesian graduation is to set forth the smoothness characteristics that are to be satisfied by the graduated estimates. We shall first discuss smoothness over age groups for a given calendar period. Then we shall discuss smoothness over calendar periods.

For each calendar period ℓ we shall assume in our illustration that the prior distribution on $\underline{v}_\ell = (v'_{1\ell},\ldots,v'_{K\ell})'$ is multivariate normal with mean $\underline{m}_\ell = (m'_{1\ell},\ldots,m'_{K\ell})'$ and covariance matrix

$$A = \left(\left(\frac{\rho_1^{|i-j|}}{4\sqrt{L_i}\sqrt{L_j}}\right)\right), \quad i = 1,\ldots,K,\ j = 1,\ldots,K,\quad 0 < \rho_1 < 1.$$

This form for the prior covariance matrix is selected not only because it meets the technical requirements of being symmetric and positive definite but also because each of its parameters has a fairly natural interpretation. The prior information about transformed forces of mortality comes from earlier mortality studies. The exposures generated in these earlier studies are generally known and hence serve as a starting point for determining L_i', $i = 1,\ldots,K$. In many cases the graduator may want to reduce the observed exposures if he feels setting the L_i''s equal to these exposures may overstate the certainty with which current transformed forces are known.

The parameter ρ_1 has a direct interpretation as the reduction in the variance of a transformed force if the value of an adjacent force were given. In some cases the choice of a single ρ_1 for all ages may be simplistic and not adequately represent the prior relationship among transformed forces at different ages. Two choices, each of which requires the specification of additional parameters, are available. First, we may choose a larger class of covariance matrices or attempt to specify all the elements of A directly by a pariwise consideration of prior probability statements about transformed forces, with an adjacent value known. See Kimeldorf and Jones [(1967), section III] for a paradigm for doing this.

A second alternative is to partition A into submatrices, each having the form of A but with different ρ's, so that each submatrix reflects the belief that within different age intervals the correlation coefficient between adjacent transformed forces may be different. See Hickman and Miller (1978) and Klugman (1978) for further discussion of this point. (We note that Leonard (1973, 1978) has used covariance matrices similar to A in other Bayesian smoothing contexts.)

We now define the cxc matrix

$$C = ((\rho_2^{|i-j|})) \quad i,j = 1,\ldots,c,$$

where ρ_2 is a positive fraction measuring the correlation between adjacent calendar periods. We then take the covariance matrix of the multivariate normal prior distribution of $\underline{v} = (\underline{v}_1', \ldots, \underline{v}_c')'$ to be the Kronecker product of C and A, i.e.,

$$\underset{KcxKc}{GA} = C \circledast A = \begin{bmatrix} A & \rho_2 A & \cdots & \rho_2^{c-1} A \\ \rho_2 A & A & \cdots & \rho_2^{c-2} A \\ \vdots & \vdots & & \vdots \\ \rho_2^{c-1} & \rho_2^{c-2} A & \cdots & A \end{bmatrix}$$

This matrix is a function of only the K + 2 parameters $L_1', L_2', \ldots, L_K'$, ρ_1 and ρ_2, all of which can be given concrete interpretations. Many other such structures can be designed, but we feel this one has both reasonable simplicity and flexibility to recommend it.

The specification of the vector of prior means $\underline{m}$ can be accomplished in any number of ways. One intriging possibility is to start by assuming that the prior expected forces of mortality fall on a conveniently chosen smooth curve, such as that of Makeham. Of course, the chosen prior means will be subject to the square root transformation to produce the vector $\underline{m}$. The posterior means will then be constrained toward the selected transformed curve without being required to fall upon it, as was done by Cramér and Wold and others using specific families of distributions. This contraint toward smoothness, but with the possibility of the data overriding the shape of the curve of prior means, is certainly in the spirit of Whittaker's original Bayesian based suggestion. Cornfield and Detre (1977) explicitly mentioned these ideas in the context of Bayesian clinical life table analysis.

The Posterior Distributions

The posterior mean (mode) of $\underline{v}^{(1)}$ will be interpreted as the graduated values of the transformed mortality rates that have arisen during the observation period, while the posterior mean (mode) of $\underline{v}^{(2)}$ will be interpreted as the predicted transformed mortality rates. The mean (mode) of the predictive distribution of $\underline{u}^{(2)}$, given $\underline{u}^{(1)}$, will be taken as a forecast of the raw transformed mortality rates that will be observed as we pass through the prediction period. Of course, the latter two means are equal, but their standard errors are different.

The joint distribution of the vector $[\underline{v}^{(1)'},\underline{v}^{(2)'},\underline{u}^{(1)'},\underline{u}^{(2)'}]'$ is multivariate normal with mean $[\underline{m}^{(1)'},\underline{m}^{(2)'},\underline{m}^{(1)'},\underline{m}^{(2)'}]'$, where $\underline{m}^{(1)'} = [\underline{m}_1',\ldots,\underline{m}_{c_1}']$ and $\underline{m}^{(2)'} = [\underline{m}_{c_1+1}',\ldots,\underline{m}_c']$, and covariance matrix

$$\begin{bmatrix} GA_{11} & GA_{12} & GA_{11} & GA_{12} \\ GA_{21} & GA_{22} & GA_{21} & GA_{22} \\ GA_{11} & GA_{12} & GA_{11}+B_{11} & GA_{12} \\ GA_{21} & GA_{22} & GA_{21} & GA_{22}+B_{22} \end{bmatrix} ,$$

where the GA_{ij} define a partition of GA that is compatible with the partition of B. The conditional distributions of $\underline{v}^{(1)}$, $\underline{v}^{(2)}$, and $\underline{u}^{(2)}$, given $\underline{u}^{(1)}$ follow immediately from this result. They are all multivariate normal distributions whose means and covariance matrices are given in Table 1.

Table 1. Means and covariance matrices of conditional distributions, given $\underline{u}^{(1)}$.

Random Variable	Conditional Mean	Conditional Covariance Matrix
$\underline{v}^{(1)}$	$\underline{m}^{(1)}+GA_{11}(GA_{11}+B_{11})^{-1}(\underline{u}^{(1)}-\underline{m}^{(1)})$	$GA_{11}-GA_{11}(GA_{11}+B_{11})^{-1}GA_{11}$
$\underline{v}^{(2)}$	$\underline{m}^{(2)}+GA_{21}(GA_{11}+B_{11})^{-1}(\underline{u}^{(1)}-\underline{m}^{(1)})$	$GA_{22}-GA_{21}(GA_{11}+B_{11})^{-1}GA_{12}$
$\underline{u}^{(2)}$	$\underline{m}^{(2)}+GA_{21}(GA_{11}+B_{11})^{-1}(\underline{u}^{(1)}-\underline{m}^{(1)})$	$GA_{22}+B_{22}-GA_{21}(GA_{11}+B_{11})^{-1}GA_{12}$

The covariance of $\underline{u}^{(2)}$, given $\underline{u}^{(1)}$, depends on B_{22}, which in turn depends on a specification of the exposures to be observed in the prediction period. Naturally probability statements about future forces must depend on future exposures.

The Reversal of the Transformation

The use of the variance stabilizing square root transformation has greatly simplified the distribution theory of the transformed mortality rates, but of course we wish to make statements about the untransformed rates. We mainly consider statements about the marginal distribution of a single rate. Let v denote the transformed rate under discussion and denote its marginal distribution by $N(\theta_v,\sigma_v^2)$. We wish to analyze the marginal distribution of the untransformed rate $\eta = v^2$, which is a non-central chi-square distribution with 1 degree of freedom. The distribution function of η is obtained from

$$Pr[\eta\leq y] = Pr[v^2\leq y] = \int_{-\frac{\sqrt{y}-\theta_v}{\sigma_v}}^{\frac{\sqrt{y}-\theta_v}{\sigma_v}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}dz,$$

so probability intervals are easy to obtain. Of course the idea of using the mode of η's distribution is not sensible here, but the mean, and in fact any moment can be obtained from the formula $E(\eta^k) = E(v^{2k})$. In practice, if a point estimate is needed for decision making purposes, then θ_v^2, the square of the posterior mean (mode) of v_1 will often be used, but the minimum mean square error estimate $E(\eta) = \theta_v^2 + \sigma_v^2$ perhaps deserves attention.

Probability statements about vectors of rates require calculations with multivariate distribution functions that are not easy to handle. Tedious calculations will yield moments of the untransformed rates, however.

4. MEASURES OF FIT AND SMOOTHNESS

Evidently graduation is a complex, multidemensional process, and so it is inevitable that a graduator will have difficulty assessing at a glance the "success" of the application of any formal graduation procedure. Often the graduator is not aware of the policy issues that may be decided on the basis of his work. As a result it is difficult to formulate a loss function which may be used with the posterior or predictive distributions in a formal decision theory approach to graduation and forecasting. Thus a graduator will typically perform a variety of informal "tests" on the results from a formal procedure. If the results pass the tests, they are deemed reasonable. Reasonableness is a necessary but not sufficient condition for acceptability, which depends (presumably) on the graduator's acceptance of the criteria used by the formal procedure. Within the Bayesian approach, failure to pass a "test of reasonableness" requires examination of the prior distribution and the data for incompatibilities.

In univariate graduation the quantities most commonly computed as "test statistics" are the measures of fit, $\sum_{x=1}^{K} L_x(u_x-v_x)^2/250$, and smoothness $\sum_{x=1}^{K-m} (\Delta^m v_x)^2$, where m = 2 or 3. These quantities are useful to an experienced graduator in comparing a series of graduations.

Notice that the traditional measure of fit, $\sum_{x=1}^{K} L_x(u_x-v_x)^2/250$, is nothing more than the exponent of (1), so with $\underline{v}$ replaced by a consistent estimator this measure can be interpreted as a chi-square statistic. But from a Bayesian point of view perhaps a more natural measure of fit would be the exponent of the posterior distribution with $\underline{v}$ replaced by $\underline{u}$. This can also be interpreted as a chi-square statistic and is a generalized distance between $\underline{u}$ and $\underline{v}$ in a coordinate system defined by the covariance matrix of the posterior distribution rather than the covariance matrix of the sampling distribution (1).

The measure of fit may be extended to the two-dimensional case in an obvious way, but we can think of several ways to measure smoothness in two dimensions. For example, McKay and Wilkin (1979) use the function

$$\alpha \sum_{j=1}^{c} \sum_{i=1}^{K-2} (\Delta_v^2 v_{ij})^2 + \beta \sum_{i=1}^{K} \sum_{j=1}^{c-2} (\Delta_h^2 v_{ij})^2$$

where

Δ_v stands for "vertical forward difference",

Δ_h stands for "horizontal forward difference",

α,β are constants measuring relative emphasis on vertical and horizontal smoothness.

In the context of Bayesian graduation we believe the calculation of some generalized variances is useful. The generalized variance of a nonsingular multivariate normal distribution is simply the determinant of the covariance matrix. The generalized variances of the prior and posterior distributions are measures of concentration of these distributions about their respective means. These generalized variances, along with the generalized variance of the distribution in (1), $|B|$, measure the interaction of prior opinion and data. Finally, the generalized variance of the predictive distribution of future observations is a measure of the overall precision with which predictions can be made.

5. NUMERICAL ILLUSTRATIONS

In our first numerical illustration we graduate five and predict three calendar periods using a portion of the Cramér-Wold (1935) data. Forces of mortality ($X10^3$) are estimated for each of the twelve five-year age groups between ages 30 and 90. The observation periods are the five five-year periods between 1861 and 1885. The prediction periods are the three five-year periods between 1886 and 1900. Estimated forces are assumed to be for the middle of the age group and the middle of the calendar period. Table 2 contains the Cramér-Wold data. To illustrate predictive analysis we have ignored the data in the last three calendar periods, but these data will be useful in judging the predictions from the formal procedure. In applying the procedure the exposures to risk for the prediction periods are projected assuming exposures increase by 5% per period in each age group.

We have conducted an extensive sensitivity analysis on these data, changing the input values of the prior means, the hypothetical past exposures, and the values of ρ_1 and ρ_2. (See Hickman and Miller (1979).) For illustrative purposes here we shall display only one graduation/prediction. The prior means are taken to be the square roots of the graduated values produced by Cramér and Wold, so in this sense we are constraining toward a bivariate Makeham surface (see Table 3). The hypothetical past exposures are one-third of the reported actual exposures in the calendar period 1856-1860. We took $\rho_1 = .9$ and $\rho_2 = .5$.

Tables 4, 5, and 6 display some of the output from the Bayesian graduation/prediction analysis. Table 4 shows the observed and graduated square roots of the forces of mortality in the observation period along with appropriate standard errors and generalized variances. Generally speaking, the graduated values are only slightly different from the observed values, a result we expect because the hypothetical past exposures are relatively small. The data have caused a dramatic drop in standard errors from the prior to the posterior distributions. The ability to quote standard errors for the graduated values seems to us to be a distinct contribution of the Bayesian approach.

Table 2. Male Deaths/Exposures for 12 Five-year Age Groups and for 8 Five-year Calendar Periods taken from Cramér and Wold (1935)

Calendar Period

Age Group	1861-1865	1866-1870	1871-1875	1876-1880	1881-1885	1886-1890	1891-1895	1896-1900
30-35	5165/695 808	6144/698 606	5802/660 396	5069/677 544	4926/693 847	4721/724 605	5149/756 873	5059/766 184
35-40	6207/680 014	6747/648 580	6314/638 072	5330/627 501	5007/631 630	4795/647 280	5143/673 908	5451/724 180
40-45	6655/604 158	8277/631 658	6936/594 612	6035/603 386	5851/588 496	5341/592 426	5381/609 471	5650/643 482
45-50	6989/488 385	9014/555 840	8319/575 737	6772/558 036	6733/562 518	6444/550 033	5976/557 050	6183/577 836
50-55	7236/381 662	9418/441 812	9039/500 496	8434/532 488	7862/513 702	7310/521 246	7175/513 448	7021/521 819
55-60	8006/310 485	9652/336 390	9388/389 479	9584/452 662	9968/480 417	8823/468 254	9164/478 513	8506/472 109
60-65	9470/255 413	10739/262 404	9740/286 128	10027/339 858	11360/396 312	11513/424 098	11076/416 817	10964/427 146
65-70	10740/206 882	12124/200 294	10551/209 634	10671/234 820	12204/281 456	13166/333 466	14509/358 301	13651/354 276
70-75	10435/132 644	13419/145 472	11339/143 655	10921/155 082	11964/175 473	13549/216 792	16338/259 337	17056/279 156
75-80	8105/ 65 038	10932/ 78 316	10493/ 86 444	10280/ 89 444	10938/ 98 922	11756/115 567	15011/144 193	17402/175 988
80-85	5290/ 27 141	6012/ 29 691	6521/ 34 794	7406/ 42 174	7923/ 44 239	8428/ 50 292	10103/ 60 170	12705/ 75 412
85-90	2054/ 6 866	2461 7 407	2428/ 8 320	3091/ 10 904	3728/ 13 901	3915/ 14 708	4550/ 17 114	5596/ 21 536

Table 3. Prior Parameters for Graduation and Prediction of Cramér-Wold Data

Age Group	Hypothetical Past Exposures X10^{-3}	1861-1865	1866-1870	1871-1875	1876-1880	1881-1885	1886-1890	1891-1895	1896-1900
30-35	231	3.041	2.972	2.903	2.835	2.768	2.698	2.629	2.555
35-40	210	3.212	3.134	3.061	2.992	2.924	2.856	2.789	2.721
40-45	196	3.464	3.366	3.282	3.205	3.134	3.066	3.000	2.934
45-50	187	3.848	3.718	3.607	3.513	3.431	3.358	3.291	3.228
50-55	171	4.427	4.252	4.099	3.971	3.868	3.779	3.704	3.636
55-60	160	5.246	5.032	4.832	4.657	4.511	4.394	4.298	4.218
60-65	132	6.336	6.100	5.866	5.645	5.449	5.283	5.152	5.045
65-70	93	7.693	7.482	7.235	6.990	6.755	6.541	6.356	6.210
70-75	58	9.346	9.182	8.960	8.715	8.474	8.238	8.014	7.813
75-80	32	11.395	11.243	11.066	10.846	10.621	10.406	10.187	9.966
80-85	14	13.911	13.768	13.606	13.427	13.229	13.052	12.893	12.720
85-90	4	16.956	16.854	16.699	16.536	16.371	16.221	16.130	16.073

$\rho_1 = .9$ $\rho_2 = .5$

Table 4. Square root of observed force of mortality (sampling distribution standard error)/graduated value of square root of force of mortality (posterior standard error) for observation period. Cramér-Wold data. Prior standard errors and generalized variances also shown.

Age Group	Observed (SE)/Graduated (SE) Period 1861-1865	1866-1870	1871-1875	1876-1880	1881-1885	Prior Standard Errors of Graduate Values
30-35	2.72(.019)/ 2.81(.013)	2.96(.019)/ 3.02(.013)	2.96(.020)/ 2.94(.013)	2.74(.019)/ 2.75(.013)	2.66(.019)/ 2.69(.013)	0.033
35-40	3.02(.019)/ 3.02(.012)	3.22(.019)/ 3.25(.012)	3.14(.020)/ 3.14(.011)	2.91(.020)/ 2.92(.012)	2.82(.020)/ 2.86(.012)	0.034
40-45	3.32(.020)/ 3.32(.012)	3.62(.020)/ 3.58(.012)	3.42(.021)/ 3.40(.012)	3.16(.020)/ 3.17(.013)	3.15(.021)/ 3.12(.012)	0.036
45-50	3.78(.023)/ 3.74(.013)	4.03(.021)/ 3.99(.012)	3.80(.021)/ 3.76(.012)	3.48(.021)/ 3.49(.012)	3.46(.021)/ 3.44(.013)	0.037
50-55	4.35(.026)/ 4.32(.014)	4.62(.024)/ 4.56(.013)	4.25(.022)/ 4.24(.013)	3.98(.022)/ 3.95(.013)	3.91(.022)/ 3.89(.013)	0.038
55-60	5.08(.028)/ 5.10(.015)	5.36(.027)/ 5.31(.015)	4.91(.025)/ 4.91(.014)	4.60(.023)/ 4.59(.014)	4.56(.023)/ 4.50(.014)	0.040
60-65	6.09(.031)/ 6.12(.016)	6.40(.031)/ 6.35(.016)	5.83(.029)/ 5.87(.016)	5.43(.027)/ 5.50(.015)	5.35(.025)/ 5.38(.015)	0.044
65-70	7.20(.035)/ 7.40(.019)	7.78(.035)/ 7.73(.019)	7.09(.034)/ 7.18(.019)	6.74(.033)/ 6.78(.018)	6.58(.030)/ 6.63(.018)	0.052
70-75	8.87(.043)/ 9.03(.025)	9.60(.041)/ 9.51(.023)	8.88(.042)/ 8.91(.023)	8.39(.040)/ 8.48(.023)	8.26(.038)/ 8.32(.022)	0.066
75-80	11.16(.062)/11.13(.034)	11.81(.057)/11.71(.032)	11.02(.054)/11.06(.031)	10.72(.053)/10.66(.030)	10.52(.050)/10.52(.030)	0.089
80-85	13.96(.096)/13.71(.054)	14.23(.092)/14.45(.052)	13.69(.085)/13.68(.049)	13.25(.077)/13.29(.047)	13.38(.075)/13.22(.046)	0.134
85-90	17.30(.191)/16.81(.116)	18.23(.184)/18.16(.109)	17.08(.173)/16.99(.106)	16.84(.151)/16.50(.099)	16.38(.134)/16.43(.095)	0.250

Generalized Variances X 10^{180}

Prior	Sampling Distribution	Posterior
0.18×10^{-15}	0.38×10^{5}	0.17×10^{-35}

Table 5 reports traditional measures of smoothness for both square roots of forces and the forces themselves. The apparently minor adjustments have resulted in substantial increases in smoothness across age groups. The data are scanty across calendar periods, but the evidence suggests little or no smoothing across these periods. This is perhaps consistent with setting $\rho_2 = .5$. We also ran the minimum mean squared error values, $(\text{mean})^2$ + variance, through the smoothness calculation and found that their smoothness measures were larger than those of the actual data.

Table 6 presents both Bayesian and sampling theory measures of "fit" for the observation period. The Bayesian measure is the exponent of the posterior marginal distribution of the component of $\underline{v}^{(1)}$ defined by the row or column under discussion with u's substituted for v's. The sampling theory measure is the usual exposure weighted sum of squares of the form $\Sigma L_x(u_x-\hat{v}_x)^2/250$, where $\hat{v}_x$ is the graduated value of u_x. The overall measures of fit are computed from all the cells in the observation period. The table also shows rankings of the column and row fits. While there are some minor differences in the rankings, the two sets tend to tell the same story.

Table 7 presents some predictions of the transformed forces of mortality. If we look at these predictions as forecasts of future observed forces, we associate with them the predictive standard errors, whereas if we look at them as estimates of the underlying population forces, we associate with them the posterior standard errors.

We also have covariances of the estimates, and these would be needed if we wanted to calculate the variance of a linear combination of estimated forces.

Table 5. Sums of squares of second/third differences of prior mean observed, and graduated forces of mortality and their square roots in the observation period. Cramér-Wold data.

	Variable					
Calendar Period	m	u	v	m^2	u^2	v^2
	Differencing Across Age Groups					
				$(X10^{-4})$	$(X10^{-4})$	$(X10^{-4})$
1861-65	1.01 /0.03	1.40/ 0.30	1.06/ 0.04	0.174/0.017	0.230/0.025	0.181/0.018
1866-70	1.05 /0.03	3.17/ 0.01	1.84/ 0.22	0.179/0.019	0.519/0.268	0.353/0.073
1871-75	1.07 /0.03	1.46/ 0.18	1.32/ 0.08	0.177/0.019	0.244/0.036	0.224/0.032
1876-80	1.10 /0.03	2.06/ 1.18	1.22/ 0.05	0.176/0.017	0.302/0.111	0.194/0.022
1881-85	1.15 /0.03	1.23/ 0.41	1.23/ 0.04	0.179/0.018	0.158/0.019	0.190/0.018
				$(X10^{-4})$	$(X10^{-4})$	$(X10^{-4})$
Average	1.08 /0.03	1.86/ 0.82	1.33/ 0.09	0.177/0.018	0.291/0.092	0.228/0.032
Age Group	Differencing Across Calendar Period					
	$(X10^3)$					
30-35	0.003/0.003	0.14/ 0.15	0.11/ 0.09	0.006/0.000	4.43/ 4.92	3.72/ 3.03
35-40	0.052/0.010	0.12/ 0.10	0.15/ 0.12	0.004/0.000	4.67/ 3.78	3.96/ 4.91
40-45	0.280/0.046	0.32/ 0.29	0.22/ 0.19	0.018/0.003	15.06/ 14.25	10.32/ 9.20
45-50	0.826/0.034	0.32/ 0.29	0.28/ 0.26	0.062/0.003	18.41/ 17.04	16.22/ 15.80
50-55	1.72 /0.016	0.44/ 0.54	0.36/ 0.38	0.169/0.001	35.44/ 44.64	27.43/ 30.33
55-60	1.65 /0.125	0.61/ 0.76	0.44/ 0.50	0.251/0.006	65.23/ 84.51	46.15/ 55.28
60-65	0.767/0.236	0.89/ 1.09	0.58/ 0.69	0.216/0.024	134.38/ 172.92	88.45/ 110.12
65-70	1.430/1.582	1.74/ 2.56	0.88/ 1.09	0.282/0.368	390.20/ 588.59	198.28/ 255.87
70-75	4.066/2.065	2.30/ 2.85	1.28/ 1.58	1.16 /0.750	785.34/1006.3	436.27/ 588.71
75-80	2.462/1.668	2.36/ 3.97	1.64/ 2.16	1.00 /0.828	1242.8 /2111.2	855.06/1154.4
80-85	1.039/0.009	0.99/ 1.05	2.54/ 3.59	0.603/0.008	757.59/ 816.08	2019.7 /2895.8
85-90	2.739/1.831	5.17/10.09	7.01/10.37	2.96/2.13	6482.6 / 12,673	8647.2 /12,976
	$(X10^3)$					
Average	1.42 /0.64	1.28/ 1.98	1.29/ 1.75	0.56/0.34	830 /1490	1020 /1510

Table 6. Bayesian and sampling theory measures of "fit" of square roots of forces of mortality in the observation period. Cramér-Wold data.

Calendar Period	Measure: Bayesian	Rank	Sampling Theory	Rank
1861-65	253	1	21.4	1
1866-70	140	2	11.0	2
1871-75	47	5	3.92	5
1876-80	92	4	5.26	4
1881-85	100	3	6.88	3
Age Group				
30-35	66	4	8.19	2
35-40	20	12	1.80	11
40-45	24	10	1.99	10
45-50	28	9	2.72	8
50-55	29	7	2.94	6
55-60	28	8	1.99	9
60-65	40	5	2.83	7
65-70	153	1	10.40	1
70-75	97	2	6.73	3
75-80	21	11	1.40	12
80-85	68	3	4.50	4
85-90	31	6	2.97	5
Overall	794	-	48.5	-

Table 7. Square root of observed force of mortality/ predicted value of square root of force of mortality (predictive standard error) (posterior standard error) for prediction period. Cramér-Wold data.

Age Group	Observed/Predicted		
30-35	2.55/ 2.66(.034)(.029)	2.61/ 2.61(.037)(.032)	2.57/ 2.54(.037)(.033)
35-40	2.72/ 2.82(.036)(.030)	2.76/ 2.77(.039)(.034)	2.74/ 2.71(.039)(.034)
40-45	3.00/ 3.06(.037)(.032)	2.97/ 3.00(.040)(.035)	2.96/ 2.93(.040)(.035)
45-50	3.42/ 3.36(.038)(.032)	3.28/ 3.29(.041)(.036)	3.27/ 3.23(.041)(.036)
50-55	3.74/ 3.79(.040)(.034)	3.74/ 3.71(.043)(.037)	3.67/ 3.64(.043)(.038)
55-60	4.34/ 4.39(.041)(.035)	4.38/ 4.30(.044)(.038)	4.24/ 4.22(.045)(.039)
60-65	5.21/ 5.25(.046)(.038)	5.15/ 5.13(.049)(.042)	5.07/ 5.04(.049)(.043)
65-70	6.28/ 6.48(.054)(.046)	6.36/ 6.32(.058)(.050)	6.21/ 6.19(.059)(.051)
70-75	7.91/ 8.16(.069)(.058)	7.94/ 7.98(.073)(.064)	7.82/ 7.79(.078)(.065)
75-80	10.09/10.36(.092)(.078)	10.20/10.16(.098)(.086)	9.94/ 9.95(.104)(.088)
80-85	12.95/13.04(.139)(.118)	12.96/12.89(.149)(.130)	12.98/12.72(.158)(.133)
85-90	16.32/16.25(.259)(.222)	16.31/16.14(.275)(.243)	16.12/16.08(.292)(.248)

Generalized Variances X 10^{108}

Prior	Sampling Distribution	Posterior	Predictive
0.14×10^{-8}	0.098	0.35×10^{-9}	0.22×10^{11}

Such would be the case if we were to interpolate a force in the table using an interpolation formula that was a linear combination of tabled forces or if we were to approximate an integral with a linear quadrature formula.

The second example is derived from data collected by the committee on mortality and morbidity under group and self-administered plans of the Society of Actuaries. The data are published in the annual reports of mortality and morbidity experience number of Transactions, Society of Actuaries. The data used in the example are for female lives retired under group annuity policies on or after normal retirement date. Therefore, the perplexing problem of heterogeneous data created by ill health early retirements is reduced. The exposures and deaths are reported in terms of number of lives. The experience was contributed by a group of large life insurance companies that issued a high proportion of the group annuity policies in the United States and Canada during the period covered by the example.

The prior means were obtained from the Ga-1951 female table (see Peterson (1952)). This table served as a reserve and premium basis for group annuities in the United States and Canada during the period covered by the example. The table was based on intercompany group annuity matured life experience for years 1946-1950, with respect to retirements on or after the normal retirement date. The basic data will be found in the 1951 Reports of Mortality and Morbidity Experience, Transactions, Society of Actuaries. These data were used in specifying past exposures in the prior distribution.

The prior means for the four calendar years 1953, 1958, 1963, and 1968 used in the example were obtained by applying projection scale C to the 1951 mortality probabilities as specified in the 1951 Ga-female table. Projection scale C was one of three sets of annual rates of decrease in mortality probabilities that were developed by U.S. actuaries around 1950. Scales A and B were proposed by Jenkins and Lew (1949). Scale A assumed a continuation of long term mortality decrease as had been observed during the first half of the twentieth century. Scale B assumed that after 1950 rates of mortality decrease would be smaller at ages below 60 and somewhat higher above age 60. This was based on the proposition that mortality at higher ages was most susceptible to efforts to control cardiovascular-renal diseases and cancer. Peterson's (1950) projection scale C assumed still larger rates of mortality decrease at higher ages. Table 8 compares the three projection scales for the ages used in the example.

Table 8

Average Rates of Decrease per year

Age	Scale A	Scale B	Scale C
50	.016	.0125	.0125
60	.012	.0120	.0125
65	.010	.0110	.0125
70	.008	.0095	.0125
75	.006	.0075	.0100
80	.004	.0050	.0067
85	.002	.0025	.0033
90	.000	.0000	.0000

Almost thirty years have passed since these projection scales were developed. With the benefit of hindsight one may observe that changes in mortality have not followed any of the projection scales. However, the scales were proposed only after systematic study that followed closely the outline suggested for specifying the mean of a prior distribution. That is, past mortality experience was studied and informed opinion was elicited. The work of Jenkins and Lew was exhaustive. Peterson, who built on the Jenkins and Lew foundations, had the advantage of three additional years of mortality experience and development of scientific opinion in forming projection scale C.

In projecting exposure for the predictive distribution for 1968, two methods were used. For age groups at age 71 and above, the exposure for 1963 was multiplied by the probability of survival for five years, for the central age of the group, according to the Ga-1951 female table. For age groups age 65 and below, an average growth rate of exposure for the previous two five year periods was applied to the exposure in 1963. A mixed method was used for the age 66-70 group. That is, an estimate of the expected survivors from those exposed at ages 61-65 in 1963 was obtained by the same method used for higher ages. To this was added an estimate of the increase in exposure due to new retirements based on the method used at younger ages.

Table 9 displays the actual deaths and exposures for the four years 1953, 1958, 1963, and 1968. The year 1968 is taken as the prediction year, while the other three are observation years. Table 10 displays the prior parameters and projected exposures needed for this illustration. Table 11 displays the observed and graduated transformed forces and their standard errors, while

Tables 12 and 13 display measures of smoothness and fit. Table 14 presents a comparison of several predictions of the transformed and untransformed raw force of mortality. Comparison of columns (3) and (4) and columns (7), (8), and (9) shows that passing the data from the observation period through our Bayesian procedure yields better predictions than does using the prior mean (projection scale C). Columns (5) and (10) show the results of a naive use of the data. For each age group the ratios

$$\frac{\text{rate in 1958}}{\text{rate in 1953}}$$

and

$$\frac{\text{rate in 1963}}{\text{rate in 1958}}$$

were formed the geometric mean of these ratios was applied to the 1963 rates to obtain the projection for 1968. To obtain column (5) this procedure was applied to the transformed rates, column (6) the untransformed rates. Clearly these naive projections are inferior to the Bayesian predictions.

Table 9. Female annuitant deaths/exposures for 9 five-year age groups and for 4 years. (Sources in text)

Age Group	Year 1953	1958	1963	1968
51-55	0/ 171	2/ 214	3/ 328	3/ 439
56-60	15/1371	8/ 1874	20/ 2879	28/ 3597
61-65	63/4899	87/ 7939	132/11,230	174/16,530
66-70	111/6596	235/14,463	430/22,500	529/33,360
71-75	69/2414	180/ 6451	407/13,668	611/22,109
76-80	69/ 925	115/ 2029	340/ 5387	611/11,689
81-85	35/ 269	59/ 631	158 1448	351/ 3941
86-90	9/ 62	23/ 130	59/ 363	130/ 802
91-95	2/ 10	7/ 24	13/ 62	48/ 149

Table 10. Prior parameters and predicted exposures for female annuitant data

Age Group	Hypothetical Past Exposures	Prior Means of Transformed Forces Year 1953	1958	1963	1968	Projected Exposures for 1968
51-55	577	1.95	1.89	1.83	1.78	459
56-60	4,323	2.48	2.40	2.32	2.25	4,147
61-65	12,277	3.29	3.18	3.08	2.99	16,846
66-70	13,024	4.23	4.10	3.97	3.85	33,745
71-75	5,474	5.88	5.72	5.56	5.41	20,024
76-80	2,027	7.87	7.71	7.55	7.40	10,877
81-85	519	9.99	9.87	9.75	9.63	3,658
86-90	131	12.31	12.27	12.22	12.21	795
91-95	13	15.17	15.17	15.17	15.17	147

$\rho_1 = .9$ $\rho_2 = .5$

Table 11. Square root of observed force of mortality (sampling distribution standard error)/graduated value of square root of force of mortality (posterior standard error) for observation period. Female annuitant data. Prior standard errors and generalized variances also shown.

Age Group	Observed (SE)/Graduated (SE) Year 1953	1958	1963	Prior Standard Errors of Graduated Value
51-55	0.00(1.209)/ 2.09(0.446)	3.06(1.081)/ 2.01(0.409)	3.02(0.873)/ 2.67(0.387)	0.791
56-60	3.31(0.427)/ 2.56(0.151)	2.07(0.365)/ 2.43(0.134)	2.64(0.295)/ 2.64(0.126)	0.228
61-65	3.59(0.226)/ 3.34(0.085)	3.31(0.177)/ 3.21(0.073)	3.43(0.149)/ 3.29(0.067)	0.112
66-70	4.10(0.195)/ 4.25(0.079)	4.03(0.131)/ 4.08(0.065)	4.37(0.105)/ 4.15(0.048)	0.102
71-75	5.35(0.322)/ 5.86(0.119)	5.28(0.197)/ 5.60(0.096)	5.46(0.135)/ 5.69(0.080)	0.158
76-80	8.64(0.520)/ 7.96(0.196)	7.53(0.351)/ 7.63(0.157)	7.94(0.215)/ 7.82(0.127)	0.250
81-85	11.41(0.964)/10.21(0.386)	9.67(0.629)/ 9.85(0.307)	10.45(0.416)/10.23(0.246)	0.500
86-90	12.05(2.008)/12.70(0.787)	13.30(1.387)/12.33(0.627)	12.75(0.830)/12.83(0.489)	0.913
91-95	14.14(5.000)/16.03(2.595)	17.08(3.227)/15.23(2.040)	14.48(2.008)/15.88(1.546)	2.236

Generalized Variances

Prior	Sampling Distribution	Posterior
2.6×10^{-36}	2.9×10^{-17}	$< 10^{-38}$

Table 12. Sums of squares of second/third differences of prior mean, observed, and graduated forces of mortality and their square roots in the observation period. Female annuitant data.

Variable

Year	m	u	v	m^2	u^2	v^2
	Differencing Across Age Groups					
				(X10⁻⁴)	(X10⁻⁴)	(X10⁻⁴)
1953	1.06/0.63	20.86/34.64	1.62/0.87	0.12/0.02	0.46/1.04	0.22/0.06
1958	1.07/0.58	8.89/14.84	1.08/0.50	0.13/0.02	0.37/0.22	0.12/0.01
1963	1.09/0.48	3.77/ 4.76	1.62/0.60	0.13/0.02	0.08/0.07	0.15/0.02
Average	1.07/0.56	11.17/18.08	1.44/0.66	0.13/0.02	0.30/0.44	0.16/0.03

Age Group	Differencing Across Years (only second differences computed)					
	$(X10^5)$					
51-55	0.55	9.55	0.55	0.00	91.11	11.79
56-60	1.07	3.28	0.12	0.00	87.42	2.97
61-65	2.01	0.16	0.05	0.00	7.87	2.06
66-70	3.39	0.17	0.06	0.01	11.85	3.80
71-75	4.12	0.06	0.12	0.02	6.54	16.18
76-80	2.34	2.22	0.26	0.02	569	62.13
81-85	81.4	6.32	0.55	0.29	2727	220
86-90	107	3.26	0.76	0.64	2130	478
91-95	0.00	30.63	2.14	0.00	30,157	2078
Average	$(x10^5)$ 22.43	6.18	0.51	0.11	3976.42	319.44

Table 13.. Bayesian and sampling theory measures of "fit" of square roots of forces of mortality in the observation period. Female annuitant data.

Measure

Year	Bayesian	Rank	Sampling Theory	Rank
1953	384	1	3481	1
1958	104	2	1507	3
1961	65	3	2416	2
Age Group				
51-55	41.3	2	1025	3
56-60	43.9	1	1016	4
61-65	12.1	7	617	5
66-70	21.9	4	1314	2
71-75	28.7	3	2036	1
76-80	15.3	5	529	6
81-85	12.3	6	475	7
86-90	4.2	8	152	9
91-95	2.9	9	240	8
Overall	791	-	7403	-

Table 14. Observed and Predicted Functions of Force of Mortality for Prediction Year 1968. Female Annuitant Data.

	Transformed				Untransformed				
(1) Age Group	(2) Observed u_x	(3) Prior m_x	(4) Predictive $E(v_x)^*$	(5) Naive y_x	(6) Observed u_x^2	(7) Prior m_x^2	(8) Predictive I $[E(v_x)]^2$	(9) Predictive II $[E(v_x)]^2+Var(v_x)$	(10) Naive z_x
51-55	2.614	1.78	2.19(0.95)	2.99	6.834	3.817	4.817	5.179	8.95
56-60	2.790	2.25	2.41(0.33)	2.35	7.784	6.128	5.818	5.865	8.65
61-65	3.244	2.99	3.09(0.18)	3.35	10.526	10.794	9.550	9.567	11.24
66-70	3.982	3.85	3.94(0.15)	4.51	15.857	17.904	15.487	15.502	20.37
71-75	5.257	5.41	5.47(0.22)	5.51	27.636	34.607	29.936	29.971	30.40
76-80	7.230	7.40	7.53(0.35)	7.62	52.271	61.983	56.728	56.824	58.06
81-85	9.437	9.63	9.87(0.67)	10.00	89.064	99.185	97.347	97.723	99.92
86-90	12.732	12.21	12.52(1.33)	13.10	162.095	151.41	156.64	158.13	172.29
91-95	17.948	15.17	15.53(4.09)	14.65	322.148	230.08	241.08	256.10	214.69
Predicition Error (1)		21.49**	19.00**	65.66**		0.070***	0.020***	0.018***	0.060***
(2)		-	-	-		0.080+	0.018+	0.017+	0.064+

*Predictive standard error in parentheses.

** $\sum_{x=1}^{9} L_x(u_x-\text{pred}_x)^2/250$

*** $\sum_{x=1}^{9} [L_x(u_x^2-\text{pred}_x)^2/\text{pred}_x]\times 10^{-6}$

+ $\sum_{x=1}^{9} [L_x(u_x^2-\text{pred}_x)^2/u_x^2]\times 10^{-6}$

Generalized Variances

Prior	$.16 \times 10^{-11}$
Sampling	$.38 \times 10^{-10}$
Posterior	$.72 \times 10^{-12}$
Predictive	$.45 \times 10^{-7}$

6. SUMMARY

We have presented a Bayesian model that combines graduation and prediction of mortality rates in a natural way. The prior parameters have physical interpretations that should ease their specification in many problems. The explicit statement of prior information in the model invites sensitivity analysis not only on the parameters of the prior distribution we have suggested but also on the very form of the distribution itself. The ease in interpreting the prior parameters is brought about by the use of an asymptotically correct sampling model and a variance stabilizing transformation. In actuarial and demographic studies the order of approximation implied is usually acceptable. An important product of our model is the covariance matrix of the graduated and predicted mortality rates, as well as of linear combinations of them. Finally, we remark that our approach can be applied to any set of demographic rates provided the relevant asymptotic distribiton theory is available.

7. ACKNOWLEDGEMENT

We wish to express our deep gratitude to Mr. Y.Y. Wang for superb programing assistance. We recieved a computing grant from the Graduate School of Business, University of Wisconsin, Madison.

REFERENCES

1. Brass, W. (1974). Perspectives in population prediction: illustrated by the statistics of England and Wales. J. Roy. Statist. Soc. Ser A. 137(4), 532-570 (discussion, pp. 571-583).

2. Cornfield, J. and Detre, K. (1977). Bayesian life table analysis. J. Roy. Statist. Soc., Ser. B. 39, 86-94.

3. Cox, C. and Scott, M. (1977). International studies in generation mortality. J. Inst. Actuaries, 104, 297-320 (discussion on pp. 321-333).

4. Cramér, H. and Wold (1935). Mortality variations in Sweden. A study in graduation and forecasting. Skandinavisk Aktuarietidskrift, 161-241.

5. Dickey, J. (1969). Smoothing by cheating. Ann. Math. Statist., 40, 1477-1482.

6. Hickman, J. and Miller R. (1978). Notes on Bayesian graduation. Trans. Soc. Actuaries. 29, 7-21 (discussion, pp. 23-49).

7. Hickman, J. and Miller, R. (1979). Sensitivity analysis of a Bayesian graduation/prediction model. Forthcoming.

8. Hoem, J. (1971). Point estimation of forces of transition in demographic models. J. Roy. Statist. Soc., Ser. B. 33, 275-289.

9. Hoem, J. (1972). On the statistical theory of analytic graduation. Proc., Sixth Berkeley Symp., 1, 569-600.

10. Hoem, J. (1976). The statistical theory of demographic rates. Scand. J. Statist., 3, 169-185.

11. Jenkins, W.A. and Lew, E.A. (1949). A new mortality basis for annuities, Trans. Soc. Actuaries, 1, 369-466.

12. Keyfitz, N. (1972). On future population. J. Amer. Statist. Assoc. 67, 347-363.

13. Kimeldorf, G. and Jones, D. (1967). Bayesian graduation. Trans. Soc. Actuaries. 19, 66-112.

14. Klugman, S. (1978). Discussion of 'Notes on Bayesian graduation'. Trans. Soc. Actuaries. 29, 24.

15. Leonard, T. (1973). A Bayesian method for histograms. Biometrika 60, 297-308.

16. Leonard, T. (1978). Density estimation, stochastic processes and prior information. J. Roy. Statist. Soc., Ser. B. 40, 113-132 (discussion on pp. 132-146).

17. Marquardt, D. (1974). Discussion of 'The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data.' Technometrics 16, 189-192.

18. McKay, S. and J. Wilkin (1977). Appendix: derivation of a two-dimensional Whittaker-Henderson type B graduation formula, Experience of Disabled-Worker Benefits Under OASDI, 1965-1974, U.S. Department of Health, Education, and Welfare, Actuarial Study No. 74.

19. Miller, M. (1942). Elements of Graduation. New York. Actuarial Society of America.

20. Peterson, R. (1952). Group annuity mortality. Trans. Soc. Actuaries, 7, 246-307.

21. Rao, C. (1952). Advanced Statistical Methods in Biometric Research. New York: John Wiley and Sons.

22. Schoenberg, I. (1964). Spline functions and the problem of graduation, Proceedings of the National Academy of Science (USA) 52, 947-950.

23. Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. J. Roy. Statist. Soc., Ser. B. 40, 364-372.

24. Whittaker, E. (1923). A method of graduation based on probability. Proc. Edin. Math. Soc., 41, 63-75 (see also Whittaker, E. and Robinson, G. (1944). The Calculus of Observations. London. Blackie and Sons, section 151).