

A Bayesian Adjustment of the HP Law via a Switching Nonlinear Regression Model

Dilli Bhatta and Balgobin Nandram*
Worcester Polytechnic Institute

Abstract: For many years actuaries and demographers have been doing curve fitting of age-specific mortality data. We use the eight-parameter Heligman-Pollard (HP) empirical law to fit the mortality curve. It consists of three nonlinear curves, child mortality, mid-life mortality and adult mortality. It is now well-known that the eight unknown parameters in the HP law are difficult to estimate because numerical algorithms generally do not converge when model fitting is done. We consider a novel idea to fit the three curves (nonlinear splines) separately, and then connect them smoothly at the two knots. To connect the curves smoothly, we express uncertainty about the knots because these curves do not have turning points. We have important prior information about the location of the knots, and this helps in the estimation convergence problem. Thus, the Bayesian paradigm is particularly attractive. We show the theory, method and application of our approach. We discuss estimation of the curve for English and Welsh mortality data. We also make comparisons with the recent Bayesian method.

Key words: Beta-binomial model, Gibbs sampler, median life, over parameterization, splines.

1. Introduction

Smoothing mortality curves (age-specific mortality rates) is useful to actuaries and demographers for many demographic problems. Based on the data, the observed rates are jagged or irregular along the life span or sometimes sparse in small populations. In addition, the rates for older individuals are not reliable due to age misreporting and death sparseness. So using a parametric equation to smooth the data is significant for estimating the mortality pattern and predicting the rates. We use the eight-parameter Heligman-Pollard (HP) empirical law to model mortality data across all ages. The law has been tested as a good model for fitting the pattern to many mortality data across all ages. But the

*Corresponding author.

convergence is still a very challenging problem in practice. A smooth mortality curve is desirable, and larger number of deaths does not help with the estimation problem.

For more than a century actuaries and demographers have developed several mathematical models (Gompertz, 1825; Makenham, 1860; Thiele, 1872; Siler, 1979; Heligman and Pollard, 1980; Mode and Jacobson, 1984). Heligman and Pollard (1980) proposed the most successful parametric model that is applicable across the entire age range. Heligman and Pollard (1980), Hartmann (1983), Fofar and Smith (1987), Kostaki (1992; 2000), Mode and Busby (1982), Rogers (1986) and Congdon (1993) used classical non-linear least squares procedure to fit the empirical law. However Rogers (1986) and Congdon (1993) noticed that over parameterization is a concern with the HP law and this leads to numerical instabilities. They suggested that the problem of over parameterization can be resolved by fixing the values of one or two parameters to a feasible constant.

Dellaportas, Smith and Stavropoulos (2001) adopted a Bayesian approach to overcome the problems associated with the method of non-linear least squares. They reported that with the use of informative priors the issue of over-parameterization is usually resolved; see also Sharrow and Clark (2010). There are many benefits of the Bayesian method for fitting the HP law as mentioned by Dellaportas, Smith and Stavropoulos (2001); these benefits include ease in dealing with over parameterization, non-normality, and computation. Moreover, the estimation of many quantities (useful for actuaries and demographers), such as survival probability, first to die and median lifetime, is automatic within the Bayesian approach. But there are difficulties in running their Metropolis-Hastings sampler as they needed considerable thinning in the selected chain to remove the autocorrelation. Wei, Nandram and Bhatta (2012) develop a Bayesian analysis for small areas (race-sex domains by states) to address the issues of over parameterization and sparse data in mortality curve fitting.

We make a novel adjustment to the Bayesian method. Our procedure fits the three curves (nonlinear splines) separately, and then connect them smoothly at the two knots using a switching non-linear regression model. There are two approaches to spline regression modeling (see Boor, 2001); both approaches use pieces of polynomials. Either the polynomials lie in their own domains (switching) or each polynomial starting in a domain remain in the following domains (join point regression models). Both models are quite popular in many fields such as health sciences and econometrics. Some interesting papers in join point regression models are Tiwari, Cronin, Davis, Feuer, Yu and Chib (2005), Kim, Fay, Feuer and Midthune (2000), and Ghosh, Huang, Yu and Tiwari (2009). For switching regression models, see Quandt (1958; 1960; 1972) and Chen (2007). A Bayesian approach was given by Ferreira (1975). Our review of the literature on switching

regression model has shown nothing in nonlinear regression. There are some polynomial switching regression models but zero derivatives are used at the knots. Our splines do not have turning points at the knots.

The basic eight-parameter empirical law of Heligman and Pollard (1980), known as HP first law, is given by

$$\frac{\pi_x}{1 - \pi_x} = a^{(b+x)^c} + d \exp \left\{ -\epsilon \left(\log \frac{x}{f} \right)^2 \right\} + gh^x, \quad (1)$$

where π_x is the probability that an individual, who has reached age x , will die before reaching age $x + 1$. The law expresses an odds of dying as a function of age x using three terms, representing distinct components of mortality. The first term in (1) is a rapidly declining exponential function that describes the child mortality (high in the first year of life and then decline until 10-13). The middle term, similar to a log normal function, represents the mid-life mortality. It describes the accident mortality for both males and females and maternal mortality as well. This additional mortality forms an ‘‘accident hump’’ on the natural curve. The third term is a geometrically increasing Gompertz exponential function that describes adult mortality. It generally represents the ageing or deterioration of the body (i.e., senescent mortality). It is typically assumed that the first term has no effect on the other two terms. These three terms correspond to the three curves under study. The first curve almost dies out as it enters the domains of second and third curves, thereby providing negligible contribution on these curves. This holds also for the second and third curves.

The eight parameters $0 < a, b, c, d, g < 1$, $e, h > 0$ and $f_0 \leq f \leq f_1$ (f_0 and f_1 are to be specified) have their own demographic interpretations. Here, a is approximately equal to π_1 (age 1 mortality); b is the location index of π_0 (age 0 mortality) in the interval $(\pi_1, 0.5)$; c measures the rate of mortality decline in childhood; d reflects the intensity of the hump, f indicates its location and e measures its spread; and lastly, g and h denote respectively the initial level and rate of increase of old-age mortality.

Heligman and Pollard (1980) also suggested the following eight-parameter empirical law known as HP second law

$$\pi_x = a^{(b+x)^c} + d \exp \left\{ -e \left(\log \frac{x}{f} \right)^2 \right\} + \frac{gh^x}{1 + gh^x}, \quad (2)$$

which we develop further here. Note that on the left-hand side of (1) we have $\pi_x/(1 - \pi_x)$ but on the left-hand side of (2) we have π_x . For the third term of (1) we have gh^x and the corresponding term in (2) we have $gh^x/(1 + gh^x)$. Thus, it is easy to show that each of the three terms in (2) is in $(0, 1)$. However, the sum of these three terms may not be in $(0, 1)$ but this is very unlikely.

It is worth noting that larger death counts, and therefore larger population sizes, do not help with the difficulty in estimation (i.e., nonidentifiability). We need information about the eight parameters in the HP law. Our method replaces two parameters in the HP law with two new parameters for which we have some information. These parameters are the knots where the three curves are connected; a priori we have information about the location of these knots. This motivates our Bayesian approach using the switching nonlinear regression model.

This paper develops a Bayesian methodology for fitting of the HP mortality curve. In Section 2, we present the general theory on switching regression. In Section 3, we discuss the Bayesian methodology and computation needed for estimation of the parameters. In Section 4, we apply the methodology to English and Welsh data, and we also show how to obtain median lifetimes of individuals at different ages. We also compare our model with an existing Bayesian model. Finally, Section 5 has concluding remarks.

2. Theory for Switching Regression

We want to model the three parts of HP law separately and connect them smoothly. Thus, for the switching non-linear regression model, we modify (2) as

$$\begin{aligned} \pi_x = & a^{(b+x)^c} I(x \leq \kappa_1) + d \exp \left\{ -e \left(\log \frac{x}{f} \right)^2 \right\} I(\kappa_1 < x \leq \kappa_2) \\ & + \frac{gh^x}{1 + gh^x} I(x > \kappa_2), \end{aligned} \quad (3)$$

where $\kappa_1 \leq f \leq \kappa_2$ and $I(\cdot)$ denotes the indicator function, for example, $I(x \leq \kappa_1) = 1$ if $x = 0, \dots, \kappa_1$. In (3) π_x is between zero and one because each component of the function is between zero and one with distinct domain (a feature of switching regression). In case of (2), as we discussed earlier, each part is extended to the entire age range, and by adding these together, π_x may assume values greater than unity. The switching regression model in (3) corrects this problem.

There are continuity constraints that must be incorporated to connect the three curves smoothly. At the two knots, κ_1 and κ_2 , we have

$$a^{(b+\kappa_1)^c} = d \exp \left\{ -e \left(\ln \frac{\kappa_1}{f} \right)^2 \right\}, \quad (4)$$

and

$$d \exp \left\{ -e \left(\ln \frac{\kappa_2}{f} \right)^2 \right\} = \frac{gh^{\kappa_2}}{1 + gh^{\kappa_2}}, \quad (5)$$

with $1 \leq \kappa_1 \leq f \leq \kappa_2, e > 0, h > 1$. (Note that κ_1 and κ_2 are positive.) It is worth noting that we have enforced an important condition, $h > 1$, otherwise the adult mortality curve in (3) will not be an increasing function. It appears that

this restriction is not necessary in the original HP law. Simplifying (4) and (5), we get

$$e = \frac{\ln d - (b + \kappa_1)^c \ln a}{(\ln \kappa_1 - \ln f)^2} \quad \text{and} \quad h = \left(\frac{d \exp\{-e(\ln \frac{\kappa_2}{f})^2\}}{g(1 - d \exp\{-e(\ln \frac{\kappa_2}{f})^2\})} \right)^{1/\kappa_2}. \quad (6)$$

In (3) we also incorporate the constraint,

$$d \exp\{-e(\log \frac{\kappa_1}{f})^2\} \leq d \exp\{-e(\log \frac{\kappa_2}{f})^2\}. \quad (7)$$

It is sensible to ensure that the mortality at κ_1 is at most the mortality at κ_2 . The mortality inequality (7) is an important biological condition we use in our model.

One would equate the derivative at the knots for smoothing the curve. However, in our case, there are no turning points at the knots, and, therefore, differentiation is not appropriate. We propose an interesting idea to smoothly connect the three curves at the knots by expressing uncertainty about the knots at κ_1 and κ_2 . A priori we have some information about the locations of the knots. A priori ranges of values of these parameters, the continuity constraints (4) and (5) and the constraint in (7) provide a new parameter space. For example, the constraint $\kappa_1 \leq f \leq \kappa_2$ is updated (see Theorem 1 below). Let S denote the set of constraints on $a, b, c, d, f, g, \kappa_1$ and κ_2 . In Theorem 1, important in our method, we construct S .

Theorem 1 The set of constraints, S , is given by

$$S = \{(a, b, c, d, \kappa_1, \kappa_2, g) : 0 < a, b, c, d, g < 1, d > \max(a^{(b+\kappa_1)^c}, \frac{g}{1+g}), \\ \max((\kappa_1^{\sqrt{A}} \kappa_2)^{\frac{1}{1+\sqrt{A}}}, \sqrt{\kappa_1 \kappa_2}) < f < \kappa_2; A = \frac{\ln(\frac{g}{d(1+g)})}{\ln(\frac{a^{(b+\kappa_1)^c}}{d})} > 0\}.$$

Proof of Theorem 1: From (6) we have

$$e = \frac{\ln d - (b + \kappa_1)^c \ln a}{(\ln \kappa_1 - \ln f)^2} \quad \text{and} \quad h = \left(\frac{d \exp\{-e(\ln \frac{\kappa_2}{f})^2\}}{g(1 - d \exp\{-e(\ln \frac{\kappa_2}{f})^2\})} \right)^{1/\kappa_2}.$$

Simplifying and noting that $e > 0$ and $h > 1$, we have

$$d > a^{(b+\kappa_1)^c} \quad \text{and} \quad d \exp\{-e(\ln \frac{\kappa_2}{f})^2\} > \frac{g}{(1+g)}. \quad (8)$$

Substituting e from (6) into (8) and simplifying we get

$$\left(\ln \frac{\kappa_2}{f}\right)^2 < \left(\ln \frac{\kappa_1}{f}\right)^2 \quad A \Rightarrow \left| \frac{\ln(\frac{\kappa_1}{f})}{\ln(\frac{\kappa_2}{f})} \right| > \frac{1}{\sqrt{A}}, \quad (9)$$

where $A = \ln(g/d(1+g))/\ln(a^{(b+\kappa_1)^c}/d) > 0$. The inequality (9) gives two cases,

$$\frac{\ln(\frac{\kappa_1}{f})}{\ln(\frac{\kappa_2}{f})} < -\frac{1}{\sqrt{A}} \quad \text{or} \quad \frac{\ln(\frac{\kappa_1}{f})}{\ln(\frac{\kappa_2}{f})} > \frac{1}{\sqrt{A}}.$$

The second inequality cannot be true because $A > 0$. Thus, using the first inequality we get

$$\ln f > \frac{\sqrt{A} \ln \kappa_1 + \ln \kappa_2}{1 + \sqrt{A}} \Rightarrow f > (\kappa_1^{\sqrt{A}} \kappa_2)^{\frac{1}{1+\sqrt{A}}}. \quad (10)$$

Now, because $d \exp(-e\{\ln(\kappa_1/f)\}^2) \leq d \exp(-e\{\ln(\kappa_2/f)\}^2)$ in (7), we get $|\ln(\kappa_1/f)|/\ln(\kappa_2/f) < 1$. This latter inequality gives two cases,

$$\frac{\ln(\frac{\kappa_1}{f})}{\ln(\frac{\kappa_2}{f})} < -1 \quad \text{or} \quad \frac{\ln(\frac{\kappa_1}{f})}{\ln(\frac{\kappa_2}{f})} > 1.$$

Again, the second inequality is false because $\kappa_1 < \kappa_2$. The first inequality gives

$$\ln f > \ln \sqrt{\kappa_1 \kappa_2} \Rightarrow f > \sqrt{\kappa_1 \kappa_2}. \quad (11)$$

Because $f \leq \kappa_2$, using (10) and (11), we get

$$\max\left(\frac{\sqrt{A} \ln \kappa_1 + \ln \kappa_2}{1 + \sqrt{A}}, \ln \sqrt{\kappa_1 \kappa_2}\right) < \ln f < \ln \kappa_2.$$

It follows that

$$\max\left((\kappa_1^{\sqrt{A}} \kappa_2)^{\frac{1}{1+\sqrt{A}}}, \sqrt{\kappa_1 \kappa_2}\right) < f < \kappa_2. \quad (12)$$

Because $A = \ln(g/d(1+g))/\ln(a^{(b+\kappa_1)^c}/d) > 0$ and $\ln(a^{(b+\kappa_1)^c}/d) < 0$, $\ln(g/d(1+g)) < 0$. Thus, we have

$$d > \max(a^{(b+\kappa_1)^c}, \frac{g}{1+g}). \quad (13)$$

Therefore, by (12) and (13)

$$S = \{(a, b, c, d, \kappa_1, \kappa_2, g) : 0 < a, b, c, d, g < 1, d > \max(a^{(b+\kappa_1)^c}, \frac{g}{1+g}),$$

$$\max((\kappa_1^{\sqrt{A}} \kappa_2)^{\frac{1}{1+\sqrt{A}}}, \sqrt{\kappa_1 \kappa_2}) < f < \kappa_2; A = \frac{\ln(\frac{g}{d(1+g)})}{\ln(\frac{a^{(b+\kappa_1)^c}}{d})} > 0\}.$$

Our model will lead to a likelihood function, a function of $a, b, c, d, f, g, \kappa_1$ and κ_2 which are in S ; e and h have simple formulas involving $a, b, c, d, f, g, \kappa_1, \kappa_2$. Of course, we still have eight parameters to estimate. However, we know that κ_1 and κ_2 are specified in fixed intervals of finite lengths, thereby adding an important piece of information into the model. We do not need to estimate e and h directly because they are functions of the other parameters.

Finally, we note the behavior of A in determining boundaries for f . For this purpose we use (12).

- (i) If $A \rightarrow 0$, $(\sqrt{A} \ln \kappa_1 + \ln \kappa_2)/(1 + \sqrt{A}) \rightarrow \ln \kappa_2$. Thus, $\ln \kappa_2 \leq \ln f \leq \ln \kappa_2$ which implies $\ln f = \ln \kappa_2$ (i.e. $f = \kappa_2$).
- (ii) If $A = 1$, $(\sqrt{A} \ln \kappa_1 + \ln \kappa_2)/(1 + \sqrt{A}) = \ln \sqrt{\kappa_1 \kappa_2}$. Thus, $\ln \sqrt{\kappa_1 \kappa_2} \leq \ln f \leq \ln \kappa_2$. It follows that $\sqrt{\kappa_1 \kappa_2} \leq f \leq \kappa_2$.
- (iii) As $A \rightarrow \infty$, $(\sqrt{A} \ln \kappa_1 + \ln \kappa_2)/(1 + \sqrt{A}) \rightarrow \ln \kappa_1$. Then, $\ln \sqrt{\kappa_1 \kappa_2} \leq \ln f \leq \ln \kappa_2$. Thus, as $A \rightarrow \infty$, $\sqrt{\kappa_1 \kappa_2} \leq f \leq \kappa_2$, the same as (ii).

3. Bayesian Model and Computation

In Section 3.1 we describe the model and in Section 3.2 we discuss the computation.

3.1 Bayesian Model

We study our proposed switching non-linear regression (SNR) model here and we compare it later with one existing model, the model of Dellaportas, Smith and Stavropoulos (DSS, 2001).

Let n_x be the population at risk having age in the interval $[x, x + 1)$ and y_x be the number of individuals who have reached age x and died before reaching age $x + 1$. The death counts follow binomial distribution,

$$y_x | \pi_x \stackrel{ind}{\sim} \text{Binomial}(n_x, \pi_x), \quad x = 0, \dots, t,$$

where π_x is given in (3). Let \underline{y} be the vector with elements y_x , $x = 0, \dots, t$, \underline{n} be the vector with n_x , $x = 0, \dots, t$ and $\underline{\theta} = (a, b, c, d, f, g)$. Throughout we assume that n_x , $x = 0, \dots, t$, are fixed known constants, and therefore conditioning on \underline{n} is not required a posteriori.

Then, the probability mass function of y is

$$\begin{aligned}
P(y | \underline{\theta}, \kappa_1, \kappa_2) &= \left\{ \prod_{x=0}^t \binom{n_x}{y_x} \right\} \times \left\{ \prod_{x=0}^{\kappa_1} (a^{(b+x)^c})^{y_x} (1 - a^{(b+x)^c})^{n_x - y_x} \right\} \\
&\times \left\{ \prod_{x=\kappa_1+1}^{\kappa_2} (d \exp\{-e(\ln \frac{x}{f})^2\})^{y_x} (1 - d \exp\{-e(\ln \frac{x}{f})^2\})^{n_x - y_x} \right\} \\
&\times \left\{ \prod_{x=\kappa_2+1}^t \left(\frac{gh^x}{1+gh^x} \right)^{y_x} \left(1 - \frac{gh^x}{1+gh^x} \right)^{n_x - y_x} \right\}, \\
(\underline{\theta}, \kappa_1, \kappa_2) &\in S, \quad y_x = 0, \dots, n_x, \quad x = 0, \dots, t,
\end{aligned} \tag{14}$$

where

$$e = \frac{-\ln\left(\frac{a^{(b+\kappa_1)^c}}{d}\right)}{\left(\ln \frac{\kappa_1}{f}\right)^2} \quad \text{and} \quad h = \left(\frac{d \exp\{-e(\ln \frac{\kappa_2}{f})^2\}}{g(1 - d \exp\{-e(\ln \frac{\kappa_2}{f})^2\})} \right)^{1/\kappa_2}.$$

A priori we assume the following distributions on the parameters and the knots

$$a, b, c, d, g \stackrel{iid}{\sim} \mathcal{U}(0, 1), \quad f | \kappa_1, \kappa_2 \sim \mathcal{U}(\kappa_1, \kappa_2), \quad \kappa_1 \sim \mathcal{U}(a_1, b_1), \quad \kappa_2 \sim \mathcal{U}(a_2, b_2),$$

where (a_1, b_1) and (a_2, b_2) are to be specified. These specifications will vary with the applications. Here a, b, c, d, g and f are continuous and κ_1 and κ_2 are discrete. Note again that prior specifications are not needed for e and h because these are functions of other parameters as shown in (6). Thus, the joint prior density is

$$\pi(\underline{\theta}, \kappa_1, \kappa_2) = \pi(a, b, c, d, g) \pi(f | \kappa_1, \kappa_2) \pi(\kappa_1, \kappa_2),$$

which clearly is proper. A priori the ranges of the parameters will change when the prior is incorporated in the likelihood; see Theorem 1.

Using Bayes' Theorem and Theorem 1, we get the joint posterior distribution,

$$\begin{aligned}
\pi(\underline{\theta}, \kappa_1, \kappa_2 | y) &\propto P(y | \underline{\theta}, \kappa_1, \kappa_2) \pi(\underline{\theta}, \kappa_1, \kappa_2) \\
&\propto \left\{ \prod_{x=0}^{\kappa_1} (a^{(b+x)^c})^{y_x} (1 - a^{(b+x)^c})^{n_x - y_x} \right\} \\
&\times \left\{ \prod_{x=(\kappa_1+1)}^{\kappa_2} (d \exp\{-e(\ln \frac{x}{f})^2\})^{y_x} (1 - d \exp\{-e(\ln \frac{x}{f})^2\})^{n_x - y_x} \right\} \\
&\times \left\{ \prod_{x=(\kappa_2+1)}^t \frac{(gh^x)^{y_x}}{(1+gh^x)^{n_x}} \right\} \frac{1}{(\kappa_2 - \kappa_1)}; \quad \underline{\theta}, \kappa_1, \kappa_2 \in S.
\end{aligned} \tag{15}$$

It is worth noting that because $\pi(\underline{\theta}, \kappa_1, \kappa_2)$ is proper, $\pi(\underline{\theta}, \kappa_1, \kappa_2 | \underline{y})$ is also proper. The joint posterior distribution here is complex and so it is difficult to sample from it directly. We apply the Gibbs sampler (Gelfand and Smith, 1990) which needs the conditional distribution of each parameter given the others and \underline{y} . Note that the terms in the density are not independent, rather they are linked to each other because e and h are functions of $\underline{\theta}$, κ_1 and κ_2 ; see (6).

3.2 Computation

As discussed above, we need to keep all terms each time we evaluate the function to sample the parameter from its conditional posterior density.

Letting $\underline{\theta}_{(i)}$ denote the vector of $\underline{\theta}$ excluding the i^{th} one (e.g., $\underline{\theta}_{(1)} = (b, c, d, f, g)$ and $\theta_1 = a$), the conditional posterior density of the i^{th} parameter is

$$\begin{aligned} \pi(\theta_i | \underline{\theta}_{(i)}, \kappa_1, \kappa_2, \underline{y}) &\propto \left\{ \prod_{x=0}^{\kappa_1} (a^{(b+x)^c})^{y_x} (1 - a^{(b+x)^c})^{n_x - y_x} \right\} \\ &\times \left\{ \prod_{x=(\kappa_1+1)}^{\kappa_2} (d \exp\{-e(\ln \frac{x}{f})^2\})^{y_x} (1 - d \exp\{-e(\ln \frac{x}{f})^2\})^{n_x - y_x} \right\} \\ &\times \left\{ \prod_{x=(\kappa_2+1)}^t \frac{(gh^x)^{y_x}}{(1 + gh^x)^{n_x}} \right\}, \end{aligned} \quad (16)$$

where the θ_i must be restricted according to S . The marginal conditional posterior distribution of the parameters are constrained to take values in the following intervals (i) $0 < a < d^{1/(b+\kappa_1)^c}$; (ii) $\max\{0, (\ln d / \ln a)^{1/c} - \kappa_1\} < b < 1$; (iii) $\max\{0, \ln(\ln d / \ln a) / \ln(b + \kappa_1)\} < c < 1$; (iv) $\max\{a^{(b+\kappa_1)^c}, g/(1+g)\} < d < 1$; (v) $\max\{(\kappa_1^{\sqrt{A}} \kappa_2)^{1/1+\sqrt{A}}, \sqrt{\kappa_1 \kappa_2}\} < f < \kappa_2$; and (vi) $0 < g < d/(1-d)$. We discuss how to obtain these intervals in Appendix A. The conditional posterior densities cannot be written in standard forms, so we use the gridgy Gibbs sampler (Ritter and Tanner, 1992) for $\underline{\theta}$. The idea of the gridgy Gibbs sampler is to approximate a conditional posterior density by a discrete distribution and draw samples from the discrete distribution instead.

For sampling κ_1 and κ_2 , the joint conditional posterior distribution is

$$\begin{aligned} \pi(\kappa_1, \kappa_2 | \underline{\theta}, \underline{y}) &\propto \left\{ \prod_{x=0}^{\kappa_1} (a^{(b+x)^c})^{y_x} (1 - a^{(b+x)^c})^{n_x - y_x} \right\} \\ &\times \left\{ \prod_{x=\kappa_1}^{\kappa_2} (d \exp\{-e(\ln \frac{x}{f})^2\})^{y_x} (1 - d \exp\{-e(\ln \frac{x}{f})^2\})^{n_x - y_x} \right\} \\ &\times \left\{ \prod_{x=\kappa_2}^t \frac{(gh^x)^{y_x}}{(1 + gh^x)^{n_x}} \right\} \frac{1}{(\kappa_2 - \kappa_1)}, \quad (\kappa_1, \kappa_2) \in R, \end{aligned} \quad (17)$$

where

$$R = \{(\kappa_1, \kappa_2) : a_1 \leq \kappa_1 \leq b_1, a_2 \leq \kappa_2 \leq b_2, d > \max\{a^{(b+\kappa_1)^c}, \frac{g}{1+g}\}, \\ \max\{(\kappa_1^{\sqrt{A}} \kappa_2)^{\frac{1}{1+\sqrt{A}}}, \sqrt{\kappa_1 \kappa_2}\} < f < \kappa_2\}.$$

We have implemented the griddy Gibbs sampler in (16). However, the chain got stuck at (κ_1, κ_2) . This is a common problem in Bayesian computation. One may think that the reversible jump sampler (Green, 1995) might be appropriate. This is not true because we have a single model with parameter θ . We do not have different parameters for different (κ_1, κ_2) . Thus, we obtain the joint posterior probability mass function of (κ_1, κ_2) using an independent procedure. Accordingly we draw sample from joint posterior distribution as

$$\pi(\theta, \kappa_1, \kappa_2 | y) = \pi(\theta | \kappa_1, \kappa_2, y) \pi(\kappa_1, \kappa_2 | y).$$

Note that

$$\pi(\kappa_1, \kappa_2 | y) \propto \int_{\theta \in S_{\kappa_1, \kappa_2}} \pi(\theta, \kappa_1, \kappa_2 | y) d\theta, \quad (18)$$

where

$$S_{\kappa_1, \kappa_2} = \{(a, b, c, d, \kappa_1, \kappa_2, g) : 0 < a, b, c, d, g < 1, d > \max(a^{(b+\kappa_1)^c}, \frac{g}{1+g}), \\ \max((\kappa_1^{\sqrt{A}} \kappa_2)^{\frac{1}{1+\sqrt{A}}}, \sqrt{\kappa_1 \kappa_2}) < f < \kappa_2; A = \frac{\ln(\frac{g}{d(1+g)})}{\ln(\frac{a^{(b+\kappa_1)^c}}{d})} > 0\}.$$

The analytic integration in (18) is not feasible which makes it difficult to find the joint density of κ_1 and κ_2 . In this situation, we approximate the definite integral by using Monte Carlo integration. This requires randomly chosen points at which the integrand is evaluated.

To sample these random points, we need an importance function. For this, we consider a simpler and less constrained set,

$$S_{\kappa_1, \kappa_2}^* = \{\theta : 0 < a, b, c, d, g < 1, \sqrt{\kappa_1 \kappa_2} < f < \kappa_2\}.$$

It is important to note that S_{κ_1, κ_2} is a subset of S_{κ_1, κ_2}^* . First, note that if any of the inequality constraints in S_{κ_1, κ_2} is dropped, a larger set is obtained. Second, consider (18). If $\max((\kappa_1^{\sqrt{A}} \kappa_2)^{1/1+\sqrt{A}}, \sqrt{\kappa_1 \kappa_2}) = (\kappa_1^{\sqrt{A}} \kappa_2)^{1/1+\sqrt{A}}$, S_{κ_1, κ_2}^* is a bigger set. Also, if $\max((\kappa_1^{\sqrt{A}} \kappa_2)^{1/1+\sqrt{A}}, \sqrt{\kappa_1 \kappa_2}) = \sqrt{\kappa_1 \kappa_2}$, there is no issue. Then, the importance function is

$$P^*(\theta | \kappa_1, \kappa_2, y) = \frac{1}{\kappa_2 - \sqrt{\kappa_1 \kappa_2}}, \quad \theta \in S_{\kappa_1, \kappa_2}^*.$$

Note that this is a uniform distribution on S_{κ_1, κ_2}^* . Now

$$\begin{aligned} \int_{\underline{\theta} \in S_{\kappa_1, \kappa_2}} P(\underline{\theta}, \kappa_1, \kappa_2 | \underline{y}) d\underline{\theta} &= \int_{\underline{\theta} \in S_{\kappa_1, \kappa_2}^*} I(\underline{\theta} \in S_{\kappa_1, \kappa_2}) P(\underline{\theta}, \kappa_1, \kappa_2 | \underline{y}) d\underline{\theta} \\ &= \frac{1}{P^*(\underline{\theta} | \kappa_1, \kappa_2, \underline{y})} \int_{\underline{\theta} \in S_{\kappa_1, \kappa_2}^*} I(\underline{\theta} \in S_{\kappa_1, \kappa_2}) P(\underline{\theta}, \kappa_1, \kappa_2 | \underline{y}) P^*(\underline{\theta} | \kappa_1, \kappa_2, \underline{y}) d\underline{\theta}. \\ &\approx (\kappa_2 - \sqrt{\kappa_1 \kappa_2}) \left[\frac{1}{M} \sum_{h=1}^M I(\underline{\theta} \in S_{\kappa_1, \kappa_2}) P(\underline{\theta}, \kappa_1, \kappa_2 | \underline{y})^{(h)} \right] \\ &= \hat{P}(\kappa_1, \kappa_2 | \underline{y}), \end{aligned}$$

where M is large (e.g., $M \approx 1000$). Then, the integral $\pi(\kappa_1, \kappa_2 | \underline{y})$ in (18) is approximately

$$\hat{\pi}(\kappa_1, \kappa_2 | \underline{y}) \approx \frac{\hat{P}(\kappa_1, \kappa_2 | \underline{y})}{\sum_{(\kappa_1, \kappa_2)} \hat{P}(\kappa_1, \kappa_2 | \underline{y})}, \quad a_1 \leq \kappa_1 \leq b_1, \quad a_2 \leq \kappa_2 \leq b_2.$$

Thus, we have obtained an estimate of joint posterior probability mass function of κ_1 and κ_2 .

So, our procedure is implemented by first drawing samples from $\pi(\kappa_1, \kappa_2 | \underline{y})$. Then, with each (κ_1, κ_2) we run a gridy Gibbs sampler to get $\underline{\theta}$. This latter task is similar to what is described already. We repeat this procedure a large number of times.

We compare our method with an existing model, the beta-binomial model with extra variation (Dellaportas, Smith and Stavropoulos, 2001) which we denote by DSS. The death counts follow the binomial distribution,

$$y_x | \pi_x \stackrel{ind}{\sim} \text{Binomial}(n_x, \pi_x), \quad x = 0, \dots, t.$$

Dellaportas, Smith and Stavropoulos (2001) assume that the first HP law determines age-dependent quantities m_x , with

$$\frac{m_x}{1 - m_x} = a^{(x+b)^c} + d \exp\{-e(\log \frac{x}{f})^2\} + gh^x,$$

and the probabilities of death follow a beta distribution,

$$\pi_x \sim \text{Beta}\{\zeta m_x, \zeta(1 - m_x)\}.$$

Here, $E(\pi_x | \zeta, m_x) = m_x$, $V(\pi_x | \zeta, m_x) = m_x(1 - m_x)/(1 + \zeta)$ and ζ is an unknown positive quantity determining the variance of beta distribution. If $\zeta \rightarrow \infty$, π_x will be given by (1). We also assume a priori that

$$P(\zeta) = \frac{1}{(1 + \zeta)^2}, \quad \zeta \geq 0,$$

a shrinkage prior distribution for ζ . Finally, we have taken a, b, c, d, e, f, g, h to be independent with

$$a, b, c, d, (f - f_0)/(f_1 - f_0), \quad g \stackrel{iid}{\sim} \mathcal{U}(0, 1), \quad p(e) = \frac{1}{(1 + e)^2}, \quad e > 0$$

$$\text{and } p(h) = \frac{1}{(1 + h)^2}, \quad h > 0,$$

where $f_0 = 10$ and $f_1 = 115$ as in DSS.

Denoting $\tilde{\theta}^* = (a, b, c, d, e, f, g, h)$, the resulting posterior distribution is

$$p(\tilde{\theta}^*, \tilde{\pi}, \zeta | \underline{y}) \propto \left\{ \prod_{x=0}^t \binom{n_x}{y_x} \pi_x^{y_x} (1 - \pi_x)^{n_x - y_x} \times \frac{\pi_x^{\zeta m_x - 1} (1 - \pi_x)^{\zeta(1 - m_x) - 1}}{B(\zeta m_x, \zeta(1 - m_x))} \right\}$$

$$\times \frac{1}{(1 + \zeta)^2} \frac{1}{(1 + e)^2} \frac{1}{(1 + h)^2},$$

where $B(\zeta m_x, \zeta(1 - m_x))$ is the beta function and $\tilde{\pi} = (\pi_0, \pi_1, \dots, \pi_t)$.

Again we perform the computations using the gridy Gibbs sampler. A discussion of the computation is given in Appendix B. The computation for our method takes more time than DSS method.

4. A Numerical Example

To illustrate the methodology we use English and Welsh mortality data for females (1988-1992). The data are presented in Table 1. To give a better illustration of the benefit of our method, we use a second example on US mortality data, 1999-2001, obtained from National Center for Health Statistics (NCHS). For confidentiality reasons we are prohibited from presenting the data. In Section 4.1 we discuss the results of fitting both SNR and DSS model and in Section 4.2 we provide median lifetimes for individuals at different ages.

4.1 Mortality Curve

For SNR model, we obtained the joint posterior probability mass function of (κ_1, κ_2) using the procedure as described in Section 3. We have specified κ_1 and κ_2 to be $8 \leq \kappa_1 \leq 13$ and $20 \leq \kappa_2 \leq 26$. The empirical posterior distribution of (κ_1, κ_2) are presented in Table 2. The data do have some influence over κ_1 and κ_2 because the joint posterior probability mass function is not uniform.

We draw a sample values of (κ_1, κ_2) from the joint posterior distribution. Then, we run the gridy Gibbs sampler to draw a sample of a, b, c, d, f and g from their corresponding conditional posterior distributions 101 times and pick the last one. We repeat this process 1000 times and finally, we have the 1000 sample values for each of the parameters. We found convergence by using the

Table 1: English and Welsh mortality data for females (1988-1992)

<i>Age</i>	n_x	y_x	<i>Age</i>	n_x	y_x	<i>Age</i>	n_x	y_x
0	1682000	11543	25	2078400	689	50	1375500	4130
1	1666400	940	26	2084300	698	51	1365400	4564
2	1644700	538	27	2067000	712	52	1373500	5017
3	1634400	420	28	2021400	799	53	1361300	5417
4	1610000	332	29	1963000	795	54	1335900	5786
5	1581800	250	30	1903800	787	55	1313900	6567
6	1564500	254	31	1844600	935	56	1306200	7173
7	1554700	228	32	1788000	978	57	1306200	8068
8	1549800	208	33	1745500	977	58	1314600	8809
9	1544600	215	34	1714800	1131	59	1325400	10148
10	1514300	182	35	1690300	1219	60	1330600	11390
11	1482500	200	36	1671400	1270	61	1332100	12789
12	1453900	215	37	1668000	1435	62	1328200	13999
13	1436700	204	38	1684600	1516	63	1322300	15528
14	1443000	294	39	1707600	1693	64	1323000	17368
15	1496400	339	40	1755900	1905	65	1329000	19277
16	1576800	412	41	1844500	2207	66	1344200	20991
17	1670500	535	42	1837500	2517	67	1370100	23665
18	1744500	561	43	1812200	2565	68	1408200	26365
19	1822800	592	44	1777100	2918	69	1337400	27664
20	1883200	591	45	1699800	3077	70	1249500	28397
21	1930400	640	46	1563200	3119	71	1174200	29178
22	1964400	623	47	1499200	3369	72	1098800	30437
23	2015600	653	48	1453200	3677	73	1029500	32146
24	2051700	668	49	1408400	3740	74	1052400	35728

Note: The data are obtained from Dellaportas, Smith and Stavropoulos (2001).

trace plots. Also non-significant auto-correlation coefficient among iterates was found. Similarly, for DSS we execute the gridy Gibbs sampler by drawing θ^* from (B.1) and ζ from (B.2). We have transformed e, h, f and ζ to $(0, 1)$ to perform the gridy Gibbs sampler. This shows good performance (i.e., after a small “burn in” there was negligible autocorrelation even at lag one).

In Table 3 we summarize the posterior distributions of a, b, c, d, e, f, g and h using posterior means (PM), posterior standard deviations (PSD) and 95% credible intervals (CI) for both models. There are some differences between the two methods (e.g., the PM for e in DSS is 10.77 versus 3.998 in SNR). The 95% credible interval for e is very narrow under DSS.

Table 2: Estimate of the joint posterior probability mass function of the knots (κ_1, κ_2) under the SNR model

$\kappa_1 \backslash \kappa_2$	20	21	22	23
8	2.086×10^{-03}	9.950×10^{-03}	4.177×10^{-03}	4.303×10^{-03}
9	8.720×10^{-03}	1.139×10^{-02}	8.906×10^{-03}	1.133×10^{-02}
10	2.396×10^{-02}	1.814×10^{-02}	3.631×10^{-01}	4.073×10^{-02}
11	2.289×10^{-02}	2.567×10^{-02}	3.980×10^{-02}	3.530×10^{-02}
12	7.886×10^{-03}	2.689×10^{-02}	2.015×10^{-02}	1.382×10^{-02}
13	1.605×10^{-02}	2.069×10^{-02}	1.863×10^{-02}	3.077×10^{-02}
$\kappa_1 \backslash \kappa_2$	24	25	26	
8	7.505×10^{-06}	4.648×10^{-03}	7.134×10^{-03}	
9	2.301×10^{-02}	1.947×10^{-02}	4.005×10^{-02}	
10	2.837×10^{-02}	3.389×10^{-02}	5.253×10^{-02}	
11	2.092×10^{-02}	6.159×10^{-02}	4.574×10^{-02}	
12	3.676×10^{-02}	2.983×10^{-02}	3.978×10^{-02}	
13	3.213×10^{-02}	2.281×10^{-02}	4.080×10^{-02}	

Table 3: Summaries of the posterior densities of the HP parameters for the DSS and SNR models

Parameter	Model	PM	PSD	95% CI
a	DSS	7.062×10^{-04}	7.475×10^{-05}	$(5.312 \times 10^{-04}, 7.831 \times 10^{-04})$
	SNR	6.793×10^{-04}	1.221×10^{-04}	$(5.27 \times 10^{-04}, 9.824 \times 10^{-04})$
b	DSS	1.895×10^{-02}	4.844×10^{-03}	$(1.078 \times 10^{-02}, 2.682 \times 10^{-02})$
	SNR	1.377×10^{-02}	7.450×10^{-03}	$(0.2534 \times 10^{-02}, 2.711 \times 10^{-02})$
c	DSS	10.49×10^{-02}	0.6397×10^{-02}	$(9.871 \times 10^{-02}, 1.215 \times 10^{-01})$
	SNR	8.771×10^{-02}	1.169×10^{-02}	$(6.492 \times 10^{-02}, 0.110)$
d	DSS	1.723×10^{-04}	1.280×10^{-05}	$(1.484 \times 10^{-04}, 1.894 \times 10^{-04})$
	SNR	3.411×10^{-04}	7.008×10^{-05}	$(2.089 \times 10^{-04}, 4.908 \times 10^{-04})$
e	DSS	10.707	0.117	(10.514, 10.904)
	SNR	3.998	2.398	(0.894, 9.583)
f	DSS	20.0	1.28	(18.1, 21.9)
	SNR	19.853	2.073	(15.488, 23.688)
g	DSS	1.919×10^{-05}	0.5697×10^{-06}	$(1.807 \times 10^{-05}, 1.998 \times 10^{-05})$
	SNR	2.679×10^{-05}	4.698×10^{-06}	$(1.892 \times 10^{-05}, 3.687 \times 10^{-05})$
h	DSS	1.107	0.6186×10^{-03}	(1.106, 1.108)
	SNR	1.102	3.478×10^{-03}	(1.095, 1.108)

Note: The measures are posterior mean (PM), posterior standard deviation (PSD) and 95% credible interval (CI).

We get the HP curve using $M = 1000$ iterates of the parameters for both DSS and SNR models. Figure 1 shows the result of fitting the HP law. The plots show curves of observed, estimated and 95% credible bands for the logarithm of the probabilities of death across ages. The observed mortalities at different ages are within the 95% credible bands. It is clear that under both models the estimated curve provides a good fit to the observed data but the DSS model underestimates the variability, and they specify $f_0 \leq f \leq f_1$, where f_0 and f_1 are really arbitrary. We use $f_0 = 15$ and $f_1 = 110$ as in DSS. We expect the confidence bands for the DSS model to be narrower than SNR model.

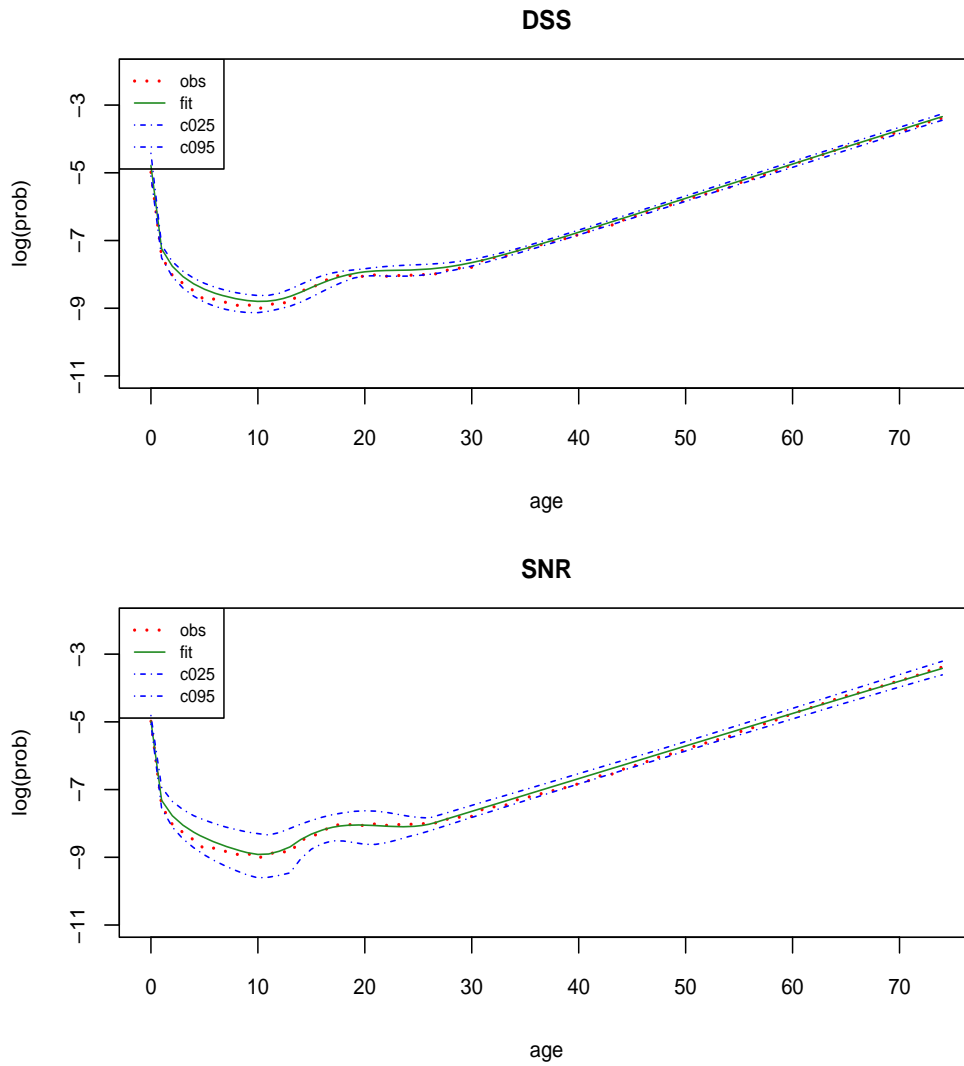


Figure 1: Observed and fitted mortality curves with 95% credible bands for the DSS and SNR models

Using box plots the empirical posterior distributions of the parameters for the two models are displayed in Figure 2. Again, larger variability for these parameters occurs in the SNR model.

In Figure 3 we have compared the fits of the DSS and SNR methods to the US mortality data. We have observed that when the DSS method is used, many of the observed mortalities are outside the 95% credible bands or very close to the bands. This is not true for the SNR method though. For age 0 to 30 the DSS bands are narrower and for age 30 to 84 wider than the SNR bands. Even when the SNR bands are narrower, they still contain the observed mortality rates. Therefore, it is reasonable to conclude that the SNR method can perform better than the DSS method.

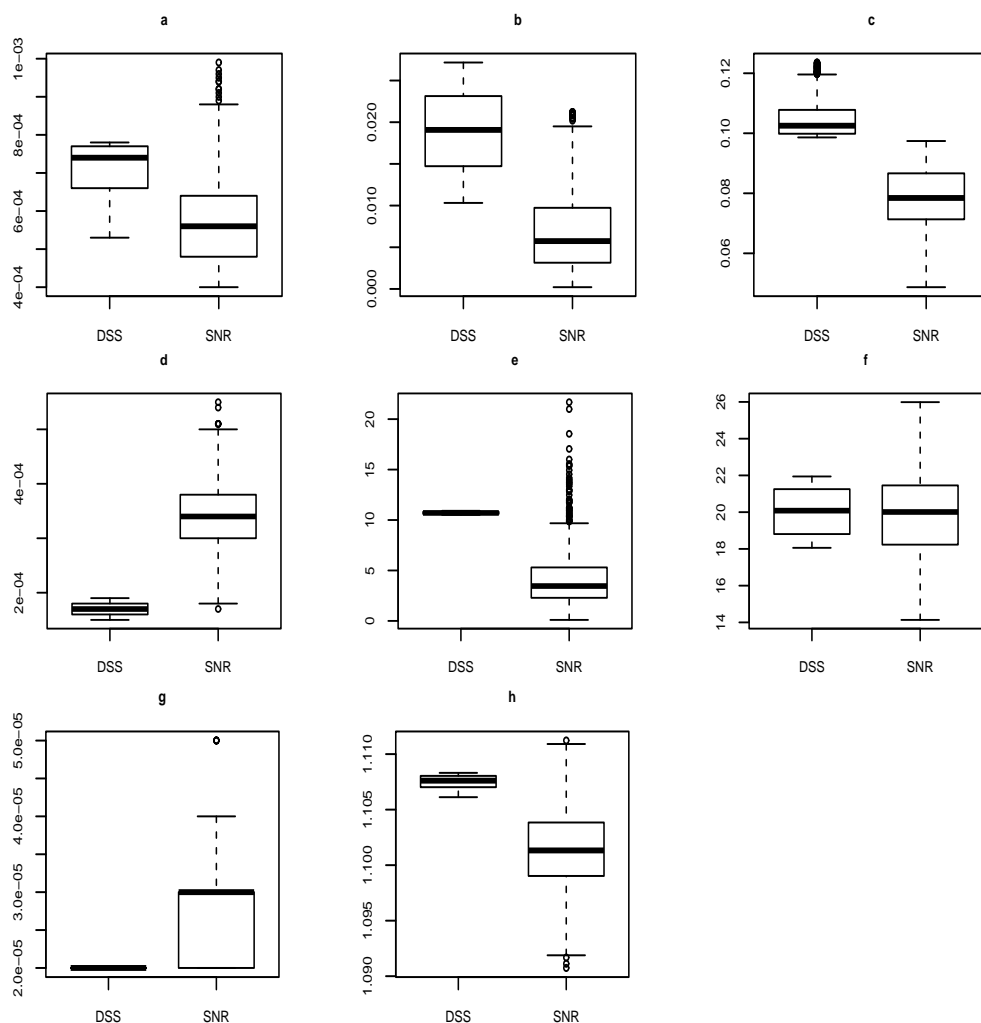


Figure 2: Box plots of the posterior simulations of parameters a, b, c, d, e, f, g and h for the DSS and SNR models

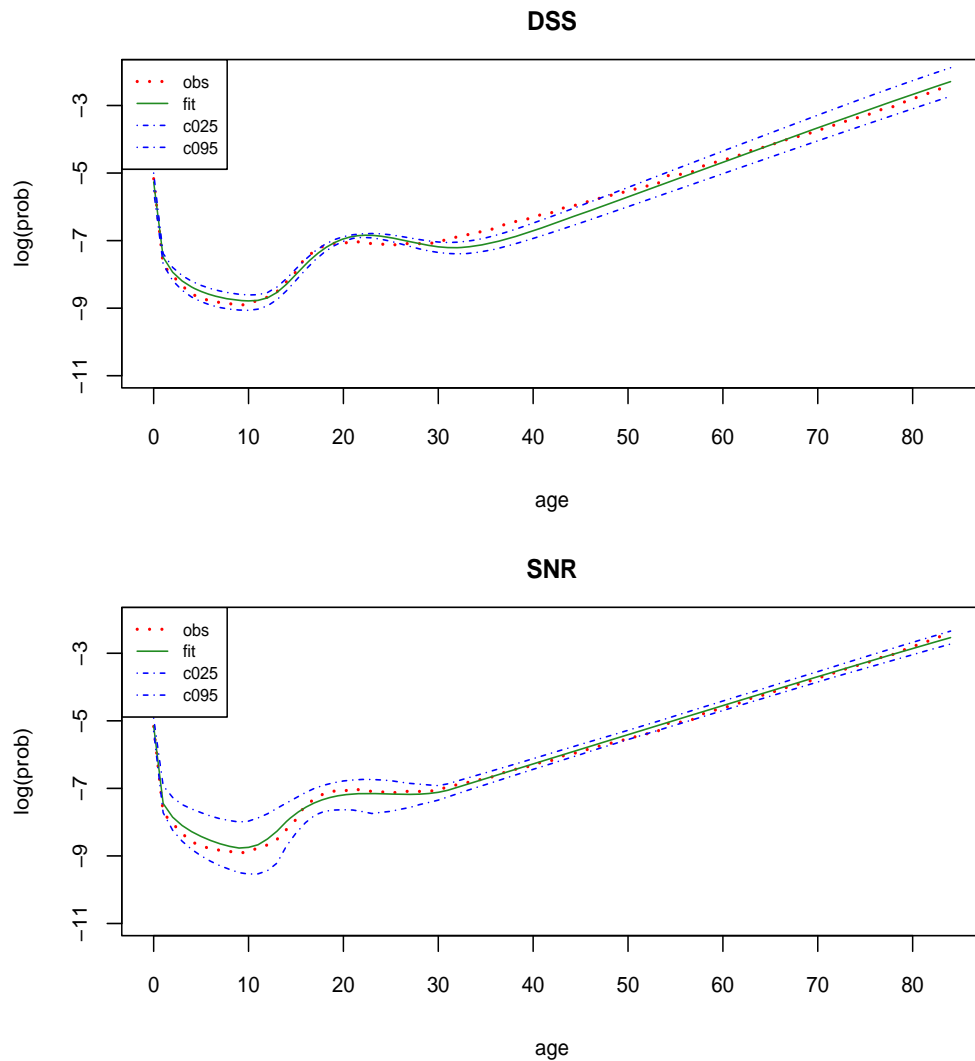


Figure 3: Observed and fitted mortality curves with 95% credible bands for the DSS and SNR models for US data

4.2 Median Lifetime

There are many useful quantities (e.g., survival probability, first to die, joint lifetime and median time to death) that are of interest to actuaries and demographers. Here we consider inference about the median lifetime, m , of an individual of age, g . It can be obtained automatically under the Bayesian method as described by DSS. The median lifetime is obtained using the survival probability

from g to m . That is, m is determined by solving the equation,

$$\prod_{x=g}^{m-1} (1 - \pi_x) = 0.5,$$

where π_x is the probabilities of death which can be obtained after fitting the HP law using the sampled parameter values of each iteration.

Because the observed deaths for the English and Welsh data are from age 0 to 74, at each iterate using the spline regression model we need to predict the probabilities beyond 74 up to 100 to allow computation of m . Then, using each sampled iterate of π_x , $x = 0, \dots, 100$, we solve the equation for m . This provides the posterior distribution of the median lifetime. The posterior inference of the median life times for adults who are at ages 55, 60, 65, 70, 75, 80 and 85 are presented in Table 4. The posterior means are similar but the intervals are narrower under the DSS method. In addition, we have obtained the posterior estimates of the median life for adults who are at ages in the range from 10 to 75. The plot of the median life is shown in Figure 4. Again, we see that the shape of the curve is same but the confidence band is narrower in DSS.

Table 4: Summaries of the posterior density of median lifetime by age

Model	55	60	65	70	75	80	85
DSS	81.4(.494) (81, 82)	81.9(.508) (81, 83)	82.4(.495) (82, 83)	83.4(.488) (83, 84)	84.9(.235) (84, 85)	87.0(0) (87, 87)	89.3(.475) (89,90)
SNR	83.2(1.42) (81, 86)	83.6(1.4) (81, 86)	84.3(1.34) (82, 87)	85.2(1.26) (83, 88)	86.6(1.18) (85, 89)	88.6(1.03) (87, 91)	91.05(0.88) (90,93)

5. Concluding Remarks

The Heligman-Pollard empirical law is very useful to fit mortality data for all ages. However, it is well known that the eight unknown parameters of the model are very difficult to estimate because there is over parameterization. While Dellaportas, Smith and Stavropoulos (2001) has pioneered Bayesian analysis for fitting the HP empirical law, we have found that it is beneficial to incorporate additional information about the eight parameters. Here, our Bayesian nonlinear spline regression model reduces the difficulty in estimating the parameters, thereby safeguarding from over parameterization. There are two difficulties which we have addressed. First, we have constructed the constraint parameter space via Theorem 1. Second, we have constructed a useful algorithm for fitting the joint posterior density.

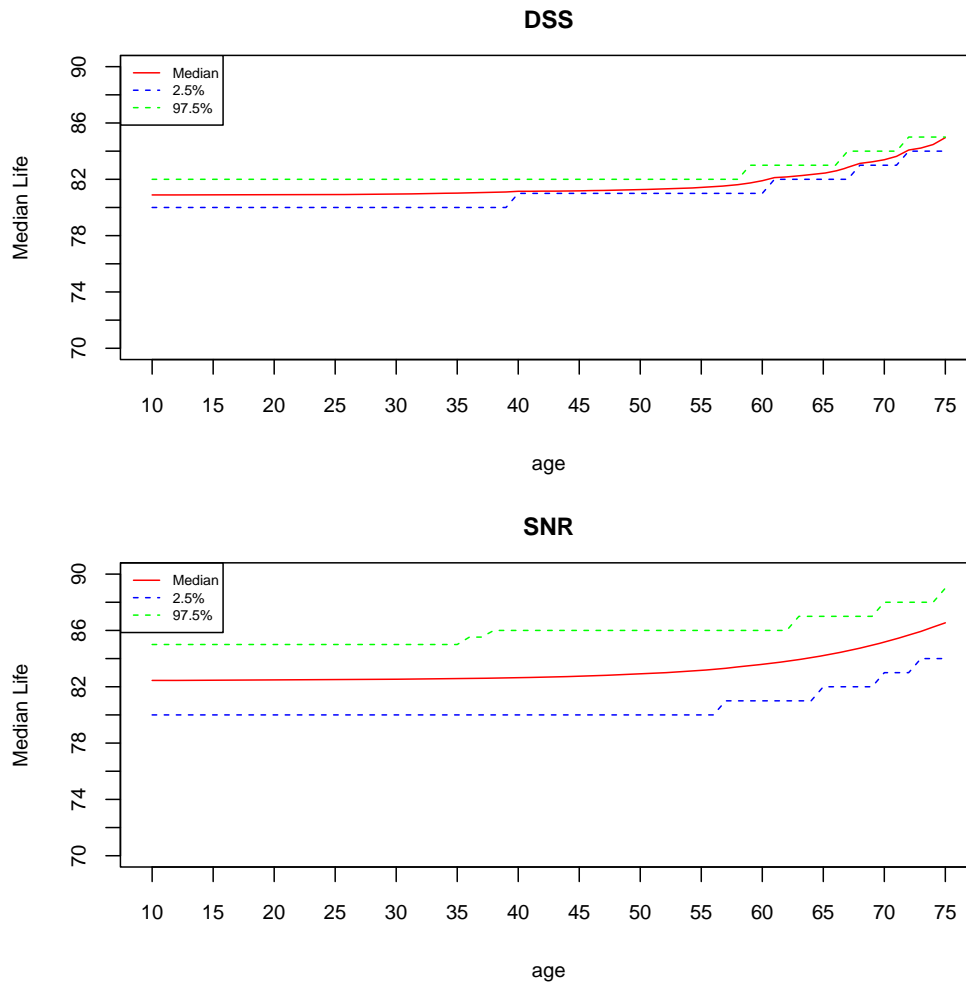


Figure 4: Expected median lifetimes with 95% credible bands of individuals, age 10 to 75, under DSS and SNR models

We fit the three curves separately and then connect them smoothly using switching regression. Like Dellaportas, Smith and Stavropoulos (2001), we have incorporated the HP law directly in the model. The results show that the estimated curves provide reasonable fits to the observed mortalities. We have also compared the results obtained from our model with Dellaportas, Smith and Stavropoulos (2001). Our switching non-linear regression model fits are similar to Dellaportas, Smith and Stavropoulos (2001) for the English and Welsh data but the confidence bands are wider. However, fitting both the model to US mortality data, 1999-2001, we found that our model fits better than one of the models of Dellaportas, Smith and Stavropoulos (2001). In addition, for both models we have obtained median lifetimes for individuals at different ages for the English

and Welsh data.

In our approach we have used the standard binomial assumption for the death counts. But both n_x and y_x are large, so we can make an accurate normal approximation via the central limit theorem. As in the Lee-Carter model (Lee and Carter, 1992), letting $\hat{m}_x = y_x/n_x$ denote the observed mortality rate, we can take

$$\hat{m}_x \stackrel{ind}{\sim} \mathcal{N}\{\pi_x, \pi_x(1 - \pi_x)/n_x\},$$

where π_x are the population mortality rates. This framework is complicated because the variance is dependent on π_x as well. We can make a variance stabilizing transformation instead to reduce the complexity by applying the arcsine transformation (Efron and Morris, 1974) to get

$$\sin^{-1}(\sqrt{\hat{m}_x}) \stackrel{ind}{\sim} \mathcal{N}\{\sin^{-1}(\sqrt{\pi_x}), 1/4n_x\}.$$

Then we can proceed as in our switching regression model. Perhaps a more efficient algorithm can be obtained in this manner.

Appendix A: Support for the Conditional Posterior Density

The supports for the conditional posterior densities of the parameters are needed to perform the gridgy Gibbs sampler. Note that $0 < a, b, c, d, g < 1$.

- (i) For a (given b, c, d and g, κ_1, κ_2), we have $d > a^{(b+\kappa_1)^c}$, and it follows that $a < \exp\{\ln d / (b + \kappa_1)^c\}$. Because $0 < a < 1$, and $\exp\{\ln d / (b + \kappa_1)^c\} < 1$, we get $0 < a < d^{1/(b+\kappa_1)^c}$.
- (ii) For b (given $a, c, d, g, \kappa_1, \kappa_2$), we have $d > a^{(b+\kappa_1)^c}$, and it follows that $b > \{\ln d / \ln a\}^{1/c} - \kappa_1$. Because $0 < b < 1$, we get $\max\{0, \{\ln d / \ln a\}^{1/c} - \kappa_1\} < b < 1$.
- (iii) For c (given $a, b, d, g, \kappa_1, \kappa_2$), we have $d > a^{(b+\kappa_1)^c}$, and it follows that $c > \ln(\ln d / \ln a) / \ln(b + \kappa_1)$. Note here that $\ln(a) < 0$ and $\ln(d) < 0$. Because $0 < c < 1$, we get $\max\{0, \ln(\ln d / \ln a) / \ln(b + \kappa_1)\} < c < 1$.
- (iv) For d (given $a, b, c, g, \kappa_1, \kappa_2$), we have $\max\{a^{(b+\kappa_1)^c}, g/(1+g)\} < d < 1$.
- (v) For f (given $a, b, c, d, g, \kappa_1, \kappa_2$), we have $\max\{(\kappa_1^{\sqrt{A}} \kappa_2)^{\frac{1}{1+\sqrt{A}}}, \sqrt{\kappa_1 \kappa_2}\} < f < \kappa_2$.
- (vi) For g (given d), we have $d > g/(1+g)$, and it follows that $g < d/(1-d)$. Because $0 < g < 1$, we get $0 < g < \min(1, d/(1-d))$.

Appendix B: Collapse Gibbs Sampler in Beta-binomial with Extra Variations

With the intention to use a collapsed Gibbs sampler, we integrate $\underline{\pi}$ from the joint posterior density to get

$$p(\underline{\theta}^*, \zeta | y) \propto \prod_{x=0}^t \binom{n_x}{y_x} \frac{B(y_x + \zeta m_x, n_x - y_x + \zeta(1 - m_x))}{B(\zeta m_x, \zeta(1 - m_x))} \frac{1}{(1 + \zeta)^2} \frac{1}{(1 + e)^2} \frac{1}{(1 + h)^2},$$

with

$$m_x = \frac{a^{(x+b)^c} + d \exp\{-e(\log \frac{x}{f})^2\} + gh^x}{1 + a^{(x+b)^c} + d \exp\{-e(\log \frac{x}{f})^2\} + gh^x},$$

where $0 < a, b, c, d, g < 1$, $e > 0$, $f_0 < f < f_1$ and $h > 1$. The conditional distributions of $\underline{\theta}^*$ and ζ are obtained as

$$p(\underline{\theta}^* | y, \zeta) \propto \prod_{x=0}^t \frac{B(y_x + \zeta m_x, n_x - y_x + \zeta(1 - m_x))}{B(\zeta m_x, \zeta(1 - m_x))} \frac{1}{(1 + \zeta)^2} \frac{1}{(1 + e)^2} \frac{1}{(1 + h)^2}, \quad (\text{B.1})$$

and

$$p(\zeta | y, \underline{\theta}^*) \propto \prod_{x=0}^t \frac{B(y_x + \zeta m_x, n_x - y_x + \zeta(1 - m_x))}{B(\zeta m_x, \zeta(1 - m_x))} \times \frac{1}{(1 + \zeta)^2}. \quad (\text{B.2})$$

Acknowledgements

We are grateful to a referee who helped to improve the manuscript. Balgobin also extends appreciation to Dr. Rong Wei, National Center for Health Statistics, for introducing this problem to him nearly a decade ago.

References

- Bacon, D. W. and Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika* **58**, 525-534.
- Boor, C. D. (2001). *A Practical Guide to Splines*. Springer, New York.
- Chen, Z. (2007). Development and inequality: evidence from an endogenous switching regression without regime separation. *Economics Letters* **96**, 269-274.

- Dellaportas, P., Smith, A. F. M. and Stavropoulos, P. (2001). Bayesian analysis of mortality data. *Journal of Royal Statistical Society, Series A* **164**, 275-291.
- Efron, B. and Morris, C. N. (1974). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association* **70**, 311-319.
- Ferreira, E. P. (1975). A Bayesian analysis of a switching regression model known number of regimes. *Journal of the American Statistical Association* **70**, 370-374.
- Forfar, D. O. and Smith, D. M. (1987). The changing shape of English life tables. *Transactions of the Faculty of Actuaries* **40**, 98-133.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398-409.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515-533.
- Ghosh, P., Huang, L., Yu, B. and Tiwari, R. C. (2009). Semiparametric Bayesian approaches to joinpoint regression for population-based cancer survival data. *Computational Statistics and Data Analysis* **53**, 4073-4082.
- Ghosh, P., Basu, S. and Tiwari, R. C. (2009). Bayesian analysis of cancer rates from SEER program using parametric and semiparametric JoinPoint regression models. *Journal of the American Statistical Association* **104**, 439-452.
- Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society* **115**, 513-585.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711-732.
- Heligman, L. and Pollard, J. H. (1980). The age pattern of mortality. *Journal of Institute of Actuaries* **107**, 49-80.
- Hartmann, M. (1987). Past and recent attempts to model mortality at all ages. *Journal of Official Statistics* **3**, 19-36.

-
- Hudson, D. J. (1966). Fitting segmented curves whose join points have to be estimated. *Journal of the American Statistical Association* **61**, 1097-1129.
- Kim, H. J., Fay, M. P., Feuer, E. J. and Midthune, D. N. (2000). Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine* **19**, 335-351.
- Kostaki, A. (1991). The Heligman-Pollard formula as a tool for expanding an abridged life table. *Journal of Official Statistics* **7**, 311-323.
- Kostaki, A. (1992). A nine parameter version of the Heligman-Pollard formula. *Mathematical Population Studies* **3**, 277-288.
- Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association* **87**, 659-671.
- Mode, C. J. and Busby, R. C. (1982). An eight-parameter model of human mortality—the single decrement case. *Bulletin of Mathematical Biology* **44**, 647-659.
- Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association* **53**, 873-880.
- Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association* **55**, 324-330.
- Quandt, R. E. (1972). A new approach to estimating switching regressions. *Journal of the American Statistical Association* **67**, 306-310.
- Ritter, C. and Tanner, M. A. (1992). Facilitating the Gibbs sampler: the Gibbs stopper and the Griddy-Gibbs sampler. *Journal of the American Statistical Association* **87**, 861-868.
- Robison, D. E. (1964). Estimates for the points of intersection of two polynomial regressions. *Journal of the American Statistical Association* **59**, 214-224.
- Sharrow, D. J. and Clark, S. J. (2010). A parametric investigation of mortality at all ages in a rural, South African population. Working Paper, Department of Sociology, University of Washington.
- Tiwari, R. C., Cronin, K. A., Davis, W., Feuer, E. J., Yu, B. and Chib, S. (2005). Bayesian model selection for join point regression with application to age-adjusted cancer rates. *Journal of the Royal statistical society, Series C* **54**, 919-939.

- Wei, R., Curtin, R. L. and Anderson, N. R. (2003). Modeling US mortality data for building life tables and further studies. In *Proceedings of the American Statistical Association, Biometrics Section*, 4458-4464. American Statistical Association, Alexandria, Virginia.
- Wei, R., Nandram, B. and Zha, W. R. (2004). A Bayesian analysis of mortality curves for small areas. In *Proceedings of the American Statistical Association, Biometrics Section*, 513-520. American Statistical Association, Alexandria, Virginia.
- Wei, R., Nandram, B. and Bhatta, D. (2012). A Bayesian analysis of US mortality curves for race-sex domains by state. Submitted.

Received May 25, 2012; accepted August 7, 2012.

Dilli Bhatta
Department of Mathematical Sciences
Worcester Polytechnic Institute
100 Institute Road, Worcester, MA 01609, USA
drb122@wpi.edu

Balgobin Nandram
Department of Mathematical Sciences
Worcester Polytechnic Institute
100 Institute Road, Worcester, MA 01609, USA
balnan@wpi.edu