Mortality projection based on the Wang transform

Piet de Jong and Claymore Marshall

Department of Actuarial Studies,
Macquarie University, NSW 2109, Australia.
Email: piet.dejong@mq.edu.au

Abstract

A new method for analysing and projecting mortality is proposed and examined. The method takes observed time series of survival quantiles, finds the corresponding z–scores in the standard normal distribution and forecasts the z-scores. The z–scores appear to follow a common simple linear progression in time and hence forecasting is straightforward. Analysis on the z–score scale offers useful insights into the way mortality evolves over time. The method and extensions are applied to Australian female mortality data to derive projections to the year 2100 in both survival probabilities and expectations of life.

Keywords: Mortality, forecasting, Wang transform.

1 Introduction

Mortality, and the likely progression of mortality, has been studied for many centuries. Since age explains a large proportion of the variation in the probability of death (or mortality), assessment of trends is usually done in terms of the probability of death at each age conditional on survival to the given age, called the age specific mortality or hazard.

Figure 1 displays the observed log–hazard (log of the central "death rate") for Australian females at each age. Each curve corresponds to one of the calendar years 1921 through to 2000. Generally higher curves correspond to the earlier years and the data is indicative of a downward trend experienced in much of the developed world. For each calendar year the log–mortalities

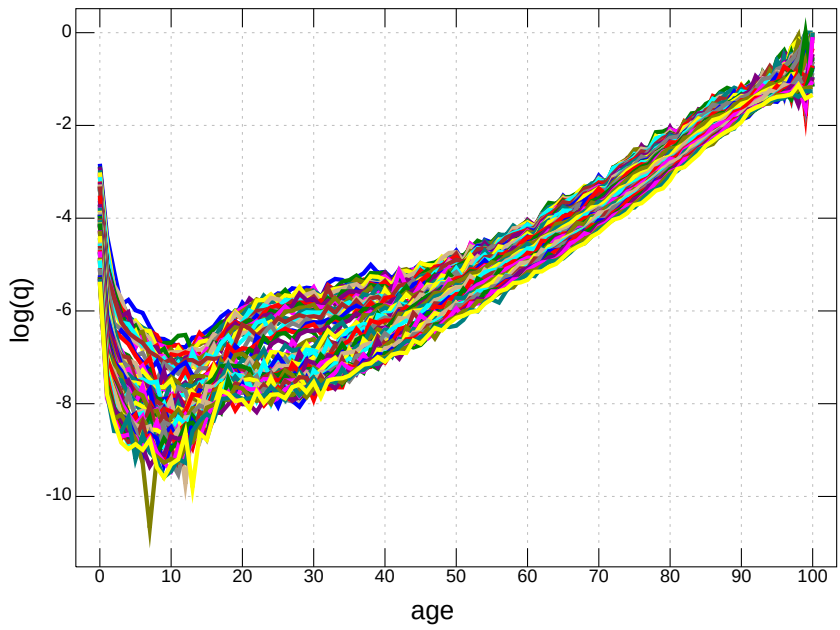

Figure 1: Log–mortality of Australian females, 1921–2000

have the characteristic shape: an initial sharp drop from age 0, followed by a slight "accident hump" (more pronounced for latter calendar years) around the late teens and early twenties, followed by an unrelenting increase into the middle and older ages. For each age, there is a discernable decrease in the log mortalities with time.

Traditional projection methods of forecasting the hazard include McNown and Rogers (1989) which uses the functional form of Heligman and Pollard (1980) to explain across age behaviour and where the across time behaviour is explained in terms of a time series model. Lee and Carter (1992) presented a model that dispenses with a parametric explanation of the age effect. In its basic form the model states that the difference between two adjacent calendar year's log–hazard is a constant λ times a function of age, independent of time. Thus λ drives the evolution of the log–hazard and, indirectly, shifts in the lifetime distribution.

A different approach to shifting arbitrary distributions is advocated in Wang (2000) in the context calculating risk adjusted insurance premiums. The approach works on the exceedence probabilities of the distribution. These are converted to z–scores having the same exceedence probabilities in the standard normal distribution. These z–scores are then uniformly shifted by an amount λ. The shifted z–scores are then transformed back, again using the standard normal, to the transformed exceedence probabilities. The

transform is called the Wang transform and depends on the single shift parameter λ. If the curve of exceedence probabilities is normal then the Wang transform corresponds to a shift λ in the mean of the distribution.

This article considers the use of the Wang transform for monitoring and forecasting mortality. The approach takes the log–hazards as in Figure 1 and converts them to survival or exceedence probabilities. These survival probabilities are converted to z–scores and it is the shifts in the z-scores between successive years which are modeled. It is shown that such shifts are virtually constant over time. It is this constancy which is exploited in the forecasting.

2 Improvements in mortality

Let q_{it} be the mortality rate in the ith year of life, $i = 1, \ldots, p$ and at time $t = 1, \ldots, n$. Thus there p ages and n calendar years. For the Australian female mortality data of Figure 2, $p = 101$ and $n = 80$. Survival probabilities, and the corresponding lifetime probabilities are defined as

$$s_{it} \equiv \prod_{j=1}^{i} (1 - q_{jt}) \ , \qquad p_{it} \equiv s_{i-1,t} - s_{it} = q_{it} s_{i,t-1} \ , \tag{1}$$

where $i = 1, \ldots, p$, $t = 1, 2, \ldots, n$ and $s_{0t} \equiv 1$. Thus s_{it} is the probability, at birth, of survival to age i, supposing the life is subject to the calendar year t mortalities $q_{1t}, \ldots, q_{pt}$. The p_{it} are the probabilities, at birth, of dying in the ith year of life. In the actuarial literature, the curve of p_{it} as a function of age i is called, somewhat ominously, the "curve of deaths."

Estimates of s_{it} and p_{it} for the Australian female mortality data of Figure 1 are displayed in the two top panels of Figure 2. Each curve represents one calendar year's data plotted against age i. Mortality improvements are indicated by the rightward shifts in the curves with time t. Clearly, modelling improvements in mortality by a linear shift independent of age in either the hazard curve, survival curve, or curve of deaths is inappropriate. For example a linear shift in the curve of deaths would imply the hazard experienced in the first year of life is translated to a higher age.

Moving from the q_{it} to the s_{it} may appear problematic since typically observations q_{it} are statistically independent across age i implying the s_{it} and z_{it} are correlated. This correlation may be difficult to deal with statistically. However with mortality forecasting the focus is on mortality progression over time. Across age dependence is of less relevance and indeed may make for a simpler statistical task in that smoothness across age is built in making for a clearer perception of time trends.

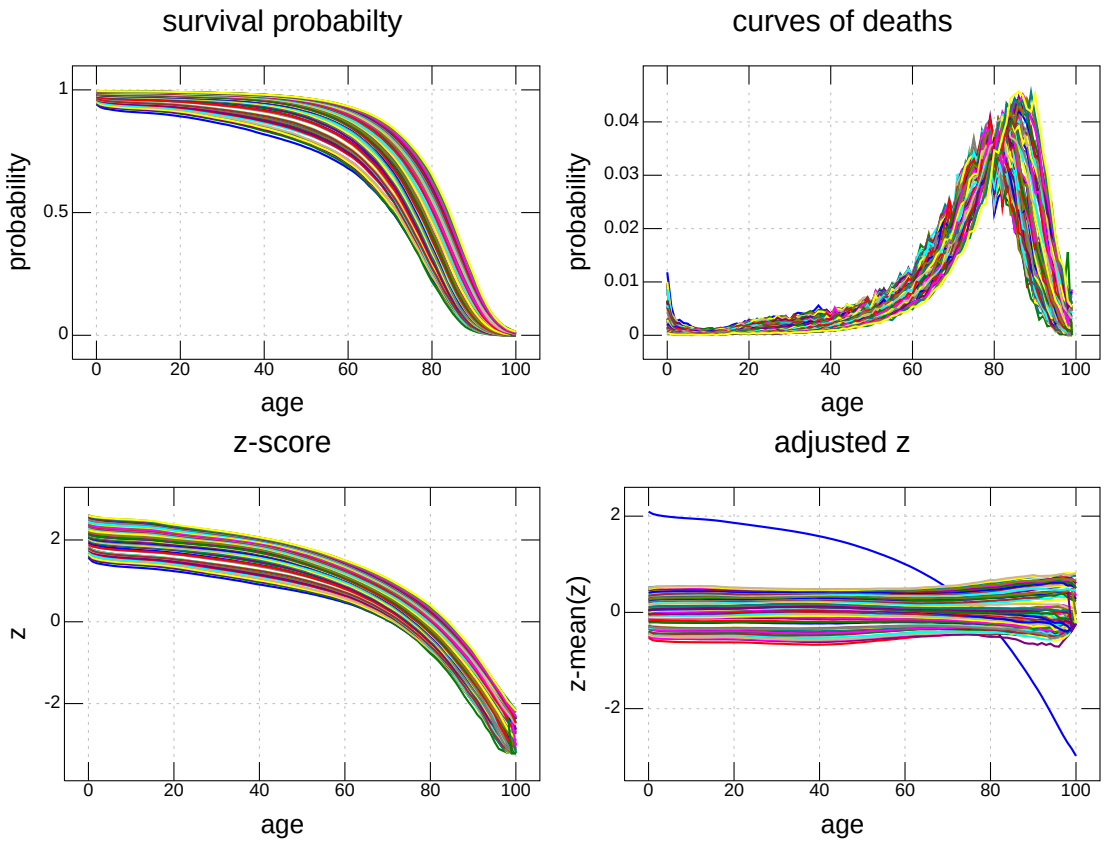

Figure 2: Survival curves and z–scores for Australian females 1921-2000.

Define z_{it} such that $s_{it} = \Phi(z_{it})$ where Φ denotes the (cumulative) standard normal distribution. Thus $z_{it} \equiv \Phi^{-1}(s_{it})$ where Φ^{-1} is the inverse mapping to Φ. The bottom two panels of Figure 2 illustrate the behaviour of mortality on the z–score scale. The lower left panel displays the relationship between z–scores and age. Each curve corresponds to a calendar year, there being $n = 80$ curves in total. The lowest curve is the year 1921, with successive curves progressing upwards representing consecutive years in the data set, up to the year 2000 which is the highest curve. Generally, the trend is for the curve for year $t+1$ to lie above the curve for year t, but this is not always the case: occasionally curves cross over. Remarkably, all the curves seem to have the same shape and, as a rough approximation, each consecutive curve from 1921 onwards appears to be a vertical shift of the previous curve. Thus the z–scores for all ages appear to grow by about the same amount over time.

The bottom right panel displays the average $\bar{z}_i = n^{-1}\sum_{t=1}^{n} z_{it}$ and the deviations from the average $z_{it} - \bar{z}_i$. Again each curve relates to a different calendar year with the lower curves corresponding to the earlier calendar years.

This article deals with the modelling and extrapolation of the z_{it} and, by implication, the s_{it} and q_{it} and related quantities. Extrapolation occurs on the time scale t. A cohort aged i at time t experiences age specific mortalities from successive time periods: q_{it}, $q_{i+1,t+1}$ and so on. It is important to realize that although the models and methods of the article project on a time scale, cohort specific results are appropriately backed out by piecing together forecasts for successive times t.

3 Analysis of mortality z–scores over time

The top left panel of Figure 3 displays $z_{it} \equiv \Phi^{-1}(s_{it})$ for the Australian female data when plotted against t. The top right panel and bottom left panel show $\hat{\lambda}_i$ and $\ln(\hat{\sigma}_i)$ for all the ages where

$$\hat{\lambda}_i \equiv \frac{\sum_{t=2}^{n}(z_{it} - z_{i,t-1})}{n-1} = \frac{z_{in} - z_{i1}}{n-1}\ , \qquad \hat{\sigma}_i^2 \equiv \frac{\sum_{t=2}^{n}\left\{(z_{it} - z_{i,t-1}) - \hat{\lambda}_i\right\}^2}{n-1}\ .$$

Thus $\hat{\lambda}_i$ is the average change over the years 1921 through to 2000 in the z–scores at each age i, while $\hat{\sigma}_i$ is the standard deviation of the changes. The average change $\hat{\lambda}_i$ is effectively constant across age although there is evidence of nonconstancy at the very young and very old ages. The log standard deviation graph indicates there is considerably more across time variation for the very old ages. This may be due to the relatively small

exposures at the higher ages and recording issues. By definition, the z–score curves never intersect on the cross-sectional scale since $z_{i-1,t} > z_{it}$, for all i and t.

The z–score age curves are distinctly linear, excluding the curves of extremely high ages (95+). Furthermore, as a rough approximation, it appears that all of the z–score age curves have approximately equal gradients over time. These broad movements in the z–scores motivate the model explained in the next section.

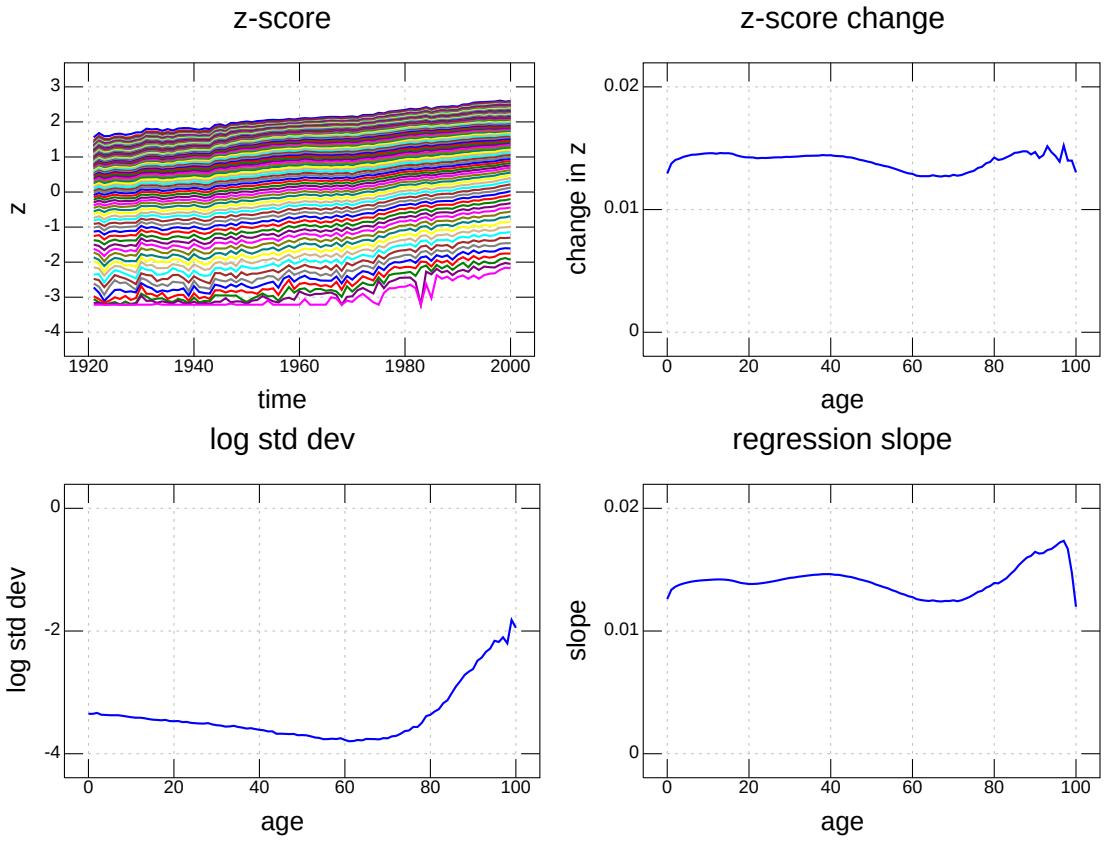

Figure 3: z–score behaviour for Australian females 1921–2000

It may be thought that the constancy across age of z–score improvement is a consequence of the Wang transform rather than an intrinsic feature of human mortality improvement. Figure 4 demonstrates that this is not the case. The figure displays, for Australian females, average z–score improvements at each age when the z–scores are computed from conditional survival curves. These conditional survival curves assume lives have reached a given middle age $m = 0, 20, 50, 65$ and 80 and then calculate the probabilities of

surviving to each higher age. The case $m = 0$ reproduces the analysis in the top right panel of Figure 3. The results for $m = 20$, $m = 50$ and so on are the graphs starting at the corresponding age. For example for ages in excess of $m = 80$ there has been a much higher increase in the z–score over the period 1921-2000 for the very high ages compared to the ages near 80. Thus a "Wang transform" analysis of the conditional distribution of remaining lifetime, conditioning on age 80, does not suggest a constant increase. To summarize, z–scores appear to linearly increase over time with a constant increase for all ages, only when z–scores are computed relative to the whole survival curve.

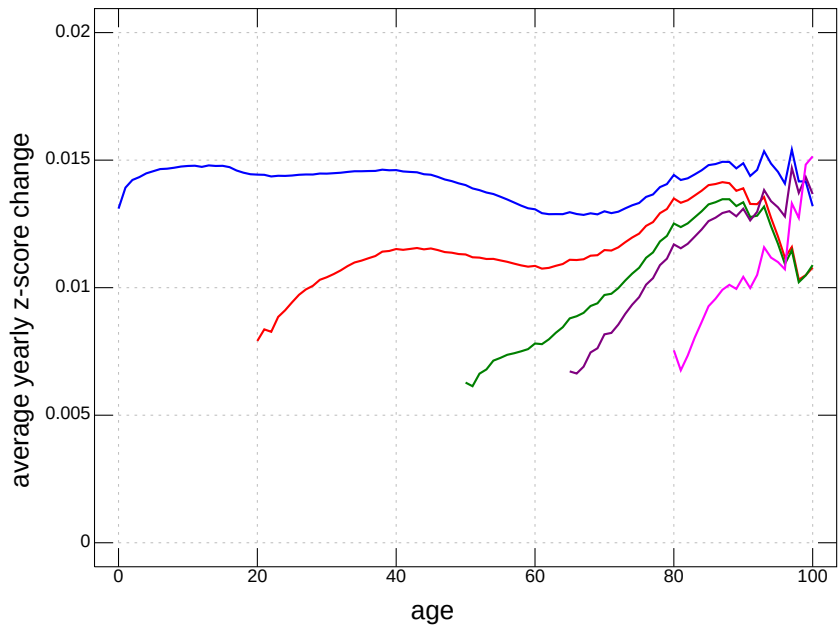

Figure 4: Conditional z–score analysis for different starting ages

4 Modelling mortality z–scores

A continuous model formalizing the notion that the z–score curve is driven by a single trend and single source of error is for $t > 0$,

$$s_{it} = \Phi(x_i'\beta + \alpha_t + \epsilon_{it}) \ , \qquad i = 1, \ldots p \ , \qquad \mathrm{d}\alpha_t = \lambda \mathrm{d}t + \sigma \mathrm{d}b_t \ , \qquad (2)$$

where $\alpha_0 = 0$, b_t is Brownian motion with zero drift and variance rate 1, λ is a superimposed constant trend and the ϵ_{it} are zero mean measurement

errors. Further similar to regression, the x_i are known regressor variables and β an unknown regression parameter vector. Equation (2) implies

$$z_{it} \equiv \Phi^{-1}(s_{it}) = x_i'\beta + \alpha_t + \epsilon_{it} \ .$$

Extensions to (2) are discussed below.

The discrete form implied by (2) stated in vector terms is

$$z_t = X\beta + 1\alpha_t + \epsilon_t \ , \qquad \alpha_{t+1} = \lambda + \alpha_t + \eta_t \ , \qquad t = 1, \ldots, n \ , \qquad (3)$$

where 1 is a vector of ones, $\eta_t = \sigma(b_{t+1} - b_t)$ is a scalar noise with mean zero and variance σ^2 and uncorrelated to the zero mean measurement error vectors ϵ_t. Components ϵ_{it} of ϵ_t are likely to be heteroskedastic. Ignoring the correlation yields the covariance matrix $\mathrm{cov}(\epsilon_t) = \sigma^2\theta\mathrm{diag}(r_1, \ldots, r_p)$ where the r_i's are relative variances. The parameter θ scales the covariance matrix of ϵ_t.

Form (3) is amenable to Kalman filtering (Harvey 1989) and hence estimation. The Kalman filter takes the z_t and transforms to serially uncorrelated increments given values of the unknown parameters λ, β and the variances of the η_t and ϵ_t. From the uncorrelated increments a normal based likelihood can be evaluated. Different parameters lead to different likelihood values and maximum likelihood estimates are located with a numerical search. Given the the maximum likelihood estimates diagnostics can be computed including estimates of the disturbances ϵ_t and η_t. Also model based projections and associated error variances are easily derived. The technology is illustrated in the next few sections.

Model (3) has similarities to the Lee and Carter (1992) model (see also De Jong and Tickle (2006)). The first equation in (3) replaces the Lee–Carter equation $\ln(m_{it}) = a_i + b_i\alpha_t + \epsilon_{it}$ where m_{it} is the vector of age specific "central" mortality rates and the a_i and b_i are age specific effects to be estimated. In the Lee–Carter setup, a single dynamic process drives mortality improvements with the differential impacts across age. These differential impacts are estimated. With (3) the mortality index manifests itself equally in all the z–scores. Differential age effects result from the transformation $s_{it} = \Phi^{-1}(z_{it})$ which in turn are converted to the q_{it}.

It may be argued that z–scores are less natural than the log–hazards $\ln q_t$ since the latter form uncorrelated estimates at each age. However the log–hazard scale magnifies features that are often irrelevant to the forecasting of broad aggregates, such as the expectation of life or proportions surviving to the various ages. For example consider the "accident" hump evident in Figure 1 between the ages of about 14 to 20. In this range the log–hazard jumps from about -10 to -8, equivalent to a jump in the mortalities from

0.00005 to 0.00034, about a 7 fold increase. While this may seem major, it is swamped by the changing pattern of mortality between the ages of say 40 and 50.

Model (3) is related to the Probit model (McCullagh and Nelder 1989). In particular suppose $\sigma = 0$ in (3). Then $s_{it} = \Phi(x_i'\beta + \lambda t + \epsilon_{it})$. Under this setup, survival probabilities are explained by nonrandom and noninteracting age and time effects.

Possible extensions to (2) or its discrete equivalent (3) include:

1. The relative variances r_i can be replaced with some smooth function of age i, for example $r_i = x_i'\gamma$. Also the r_i could incorporate a measure of exposure at the different ages.

2. The growth factor λ may be vector with different components driving rates of improvement at different ages.

3. More intricate dynamics may be appropriate. For example λ may depend on time t leading to a more complicated stochastic differential equation involving higher order differentials. For example (Arnold 1974) with the Ornstein–Uhlenbeck process $\mathrm{d}\lambda_t = \phi(\lambda_t - \mu)\mathrm{d}t + \sigma \mathrm{d}b_t$ the growth rate λ_t has, if $\phi < 0$, a tendency to return to an average value μ. In discrete time the model implies the growth rate follows an autoregressive model of order 1. Estimation and assessment is again facilitated with the Kalman filter as discussed in De Jong and Mazzi (2001).

4. The fixed age shape profile $x_i'\beta$ can be generalized to an evolving profile $x_i'\beta_t$ where β_t is a vector time series. Thus $\mathrm{d}z_{it} = x_i'\mathrm{d}\beta_t + \lambda + \sigma \mathrm{d}b_t$ and the age profile of the z–scores changes over time. A vector time series, rather than constant, description for β_t is appropriate where the z–score age shape "tilts" over time. Evidence of tilting appears in the application of the next section.

5. Particular time periods may have special characteristics. Examples include the "Spanish" flu pandemic of 1918-1920, war periods, and the years of the traffic fatality "hump." These can be regarded as interventions or shocks to the system. Dynamic models modifying (3) can be written to accommodate such shocks and the Kalman filter machinery is ideally set up to assess and quantify such shocks (De Jong and Penzer 1998). This brings out a further advantage of Wang transform: the transform molds the serial profile of mortality to a linear scale and hence well developed technology based on linearity is applicable.

5 Application to Australian female mortality data

In this application the matrix X interpolates second order b–splines (De Boor 1978) with knots at ages -0.5, 9.5, 60.5, 95, and 105, the first knot being of order 3. Thus there are 6 components in β. Maximum likelihood estimation via the Kalman filter led to the estimates of λ, σ and θ of 0.0141, 0.0261 and 232.758 respectively. The estimate of λ differs negligibly from the observed yearly average change in z–scores for each age i presented previously. This is expected since the differences $z_{it} - z_{i,t-1}$ contain all the information for λ.

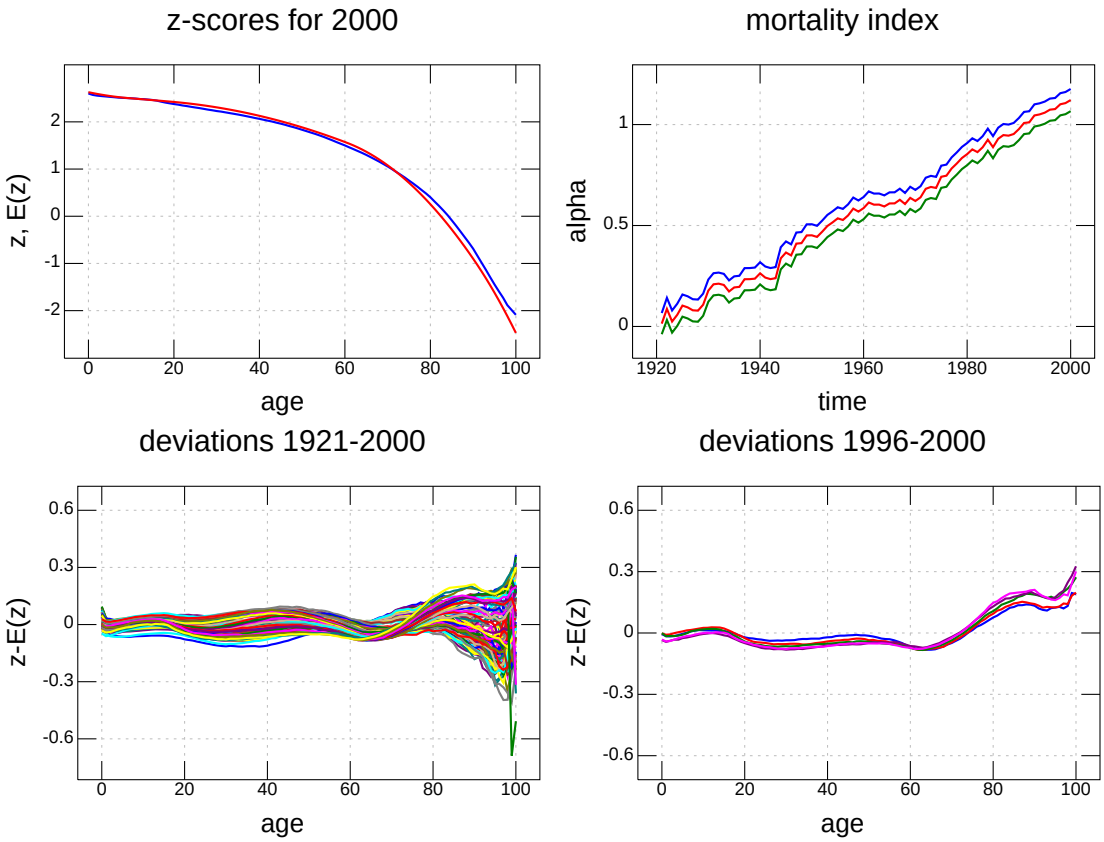

Figure 5: Estimation results for Australian female data

Figure 5 displays estimation results. The top left panel displays the actual and estimated expected z–scores for calendar year 2000. The top right panel of Figure 5 is the plot of the estimated value of α_t in each calendar year. The two outer lines represent approximate 95% confidence bands. The estimate and confidence bands are computed with the smoothing filter companion to

the Kalman filter (De Jong 1989). The graph indicates a steady improvement in mortality with some tailing off in the 1960's. The bottom left panel measures departures between fitted and actual z–scores. The curves indicate little error except at ages beyond 70. At these older ages the early decades indicate negative errors and hence actual z–scores and survival probabilities are lower than expected. For the latter decades this pattern is reversed with observed older age survival probabilities exceeding predictions. These latter years' errors are displayed in the bottom right panel of Figure 5. This systematic lack of fit at the higher ages suggest a slightly tilting age profile where mortality improves at a faster rate at the higher ages.

6 Implications for future survival probabilities and the expectation of life

In this section suppose the s_{it} and z_{it} are theoretical population constructs of which the previously defined s_{it} and z_{it} are measurements. Then $s_{it} = \Phi(z_{it})$ where $\mathrm{d}z_{it} = \lambda \mathrm{d}t + \mathrm{d}b_t$ and initial conditions $z_{i0} = x_i'\beta$. It follows from Ito's Lemma (Arnold 1974), for $i = 1, \ldots, p$,

$$\mathrm{d}s_{it} = \mathrm{d}\Phi(z_{it}) = \lambda\phi(z_{it})\mathrm{d}t + \frac{\sigma^2}{2}\frac{\mathrm{d}\phi(z_{it})}{\mathrm{d}z_{it}}\mathrm{d}t + \sigma\phi(z_{it})\mathrm{d}b_t \tag{4}$$

$$= \phi(z_{it})\left\{\left(\lambda - \frac{\sigma^2}{2}z_{it}\right)\mathrm{d}t + \sigma\mathrm{d}b_t\right\} ,$$

where ϕ denote the standard normal density. If $\sigma = 0$ then $\mathrm{d}s_{it} = \lambda\phi(z_{it})\mathrm{d}t$ and

$$s_{i,t+1} \approx \Phi(z_{it}) + \lambda\phi(z_{it}) \approx (1-\lambda)\Phi(z_{it}) + \lambda\Phi(z_{it} + \lambda) .$$

Hence improvements in survival probabilities are moderated by the normal curve. More generally, provided $\sigma = 0$,

$$s_{i,t+k} \approx \Phi(z_{it}) + k\lambda\phi(z_{it}) \approx (1-\lambda)\Phi(z_{it}) + \lambda\Phi(z_{it} + \lambda k) .$$

Significant changes in survival probabilities only possible if $\phi(z_{it})$ is not small, that is z_{it} not far from 0 or, equivalently, the survival probability is not far from 0.5. The maximum change in the survival probability occurs for age i such that $s_{it} = 0.5$, that is the median lifetime age. For the median lifetime age i, $z_{it} = 0$ and $\phi(z_{it}) = 1/\sqrt{2\pi}$. Hence s_{it} increases by about $\lambda/\sqrt{2\pi}$ per calendar year.

When $\sigma > 0$ the proportional change in the survival curve is reduced by $z_{it}\sigma^2/2$. For the early years, that is less than the median lifetime, $z_{it} > 0$ so

there is a positive reduction. For the latter years, the reduction is negative, that is the proportionate effect is greater than λ.

A similar analysis can be given for the (curtate) expected number of years, at birth, to be lived between ages 0 and $p = 100$ (Bowers, Gerber, Hickman, Jones, and Nesbitt 1997) $e_t = \sum_{i=1}^{p} s_{it}$. From (4),

$$\mathrm{d}e_t = \sum_{i=1}^{p} \mathrm{d}s_{it} = \sum_{i=1}^{p} \phi(z_{it}) \left\{ \left(\lambda - \frac{\sigma^2}{2} z_{it} \right) \mathrm{d}t + \sigma \mathrm{d}b_t \right\} = \lambda \sum_{i=1}^{p} \phi(z_{it}) \mathrm{d}t \ ,$$

where the last equality holds if $\sigma = 0$. Thus if $\lambda > 0$ the expectation of life increases. In discrete terms

$$e_{t+1} \approx e_t + \lambda \sum_{i=1}^{p} \phi(z_{it}) \ . \tag{5}$$

The effect of noise, $\sigma > 0$, is now more ambiguous. For the Australian female mortality data $\hat{\lambda} = 0.0141$, implying z–scores for all ages are increasing by about 0.0141 per annum. Figure 6 displays the behaviour of $\hat{\lambda} 1' \phi(z_t)$ as a function of t. the rate of increase in life expectancy at birth is gradually declining over time, from a value of roughly 0.3 years in 1930 to about 0.2 years in 2000. Equation (5) indicates that the size of the annual addition to life expectancy at birth depends on the magnitude of the z–scores for all ages and suggests that the annual improvement in life expectancy will continue to decline.

7 Forecasting mortality through to 2100

In this section the Australian female mortality is projected out to year 2100. Assuming $\sigma = 0$, forecast survival probabilities are

$$\hat{s}_{i,n+k} \equiv \Phi(\hat{z}_{in} + \hat{\lambda} k) \ , \qquad i = 1, \ldots, p \ , \qquad k = 1, \ldots, 100 \ , \tag{6}$$

where $n = 80$ corresponds to calendar year 2000. Here the $\hat{z}_{in}$ are the fitted or actual z–scores in 2000, sometimes called "jump off" rates.

Suppose the forecasts (6) are arranged into a $p \times 100$ matrix S with rows and columns corresponding to i and k respectively. Then

1. Row i of S corresponds to forecast survival probabilities for age i across the future calendar years through to year 2100.

2. Column k of S is the forecast "cross sectional" survival curve for calendar year $2000 + k$.

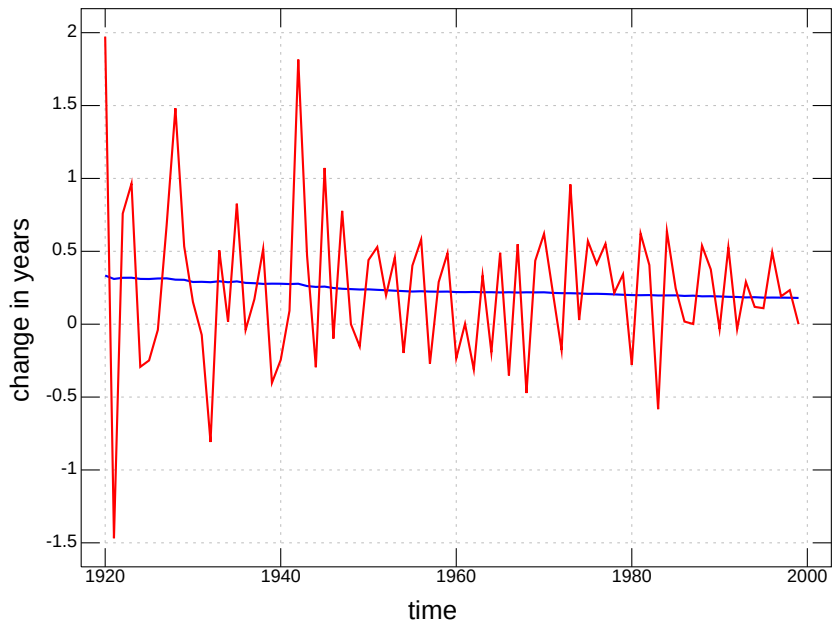

Figure 6: Observed and expected annual addition to Australian female life expectancy

3. "Cohort" survival curves are arrived by first deriving from S the corresponding death probabilities $q_{ik} = 1 - s_{i+1,k}/s_{ik}$. In turn these death probabilities are converted to survival curves as in (1). If Q is the matrix of q_{ik} then survival curves are arrived at by operating on the diagonals of Q. The main diagonal corresponds to the cohort born in 2000. Upper diagonal k corresponds to those to be born in year $2000+k$ while lower diagonal k corresponds to those aged k in 2000.

The left panel in Figure 7 displays the two forecast cross sectional survival curves for calendar years 2001 and 2100 as well as the 1921 cross sectional curve and the forecast "cohort" survival curve for those born in 2000. The bottom curve is that for 1921. The next curve up is for 2000. The highest curve is the forecast for 2100. The curve just below and barely distinguishable from the highest curve is the forecast cohort survival curve for those born in 2000. The cohort survival curve closely follows the 2100 curve since the mortality is negligible for all ages up to about 50. When they reach 50 they experience the mortality of 2050 and so on. Thus the heavier mortality is experienced once rates at the older ages have had plenty of time to improve. On the basis of present trends, more than 1 in 5 will reach age 100.

The right panel in Figure 7 displays curtate remaining life expectancies up

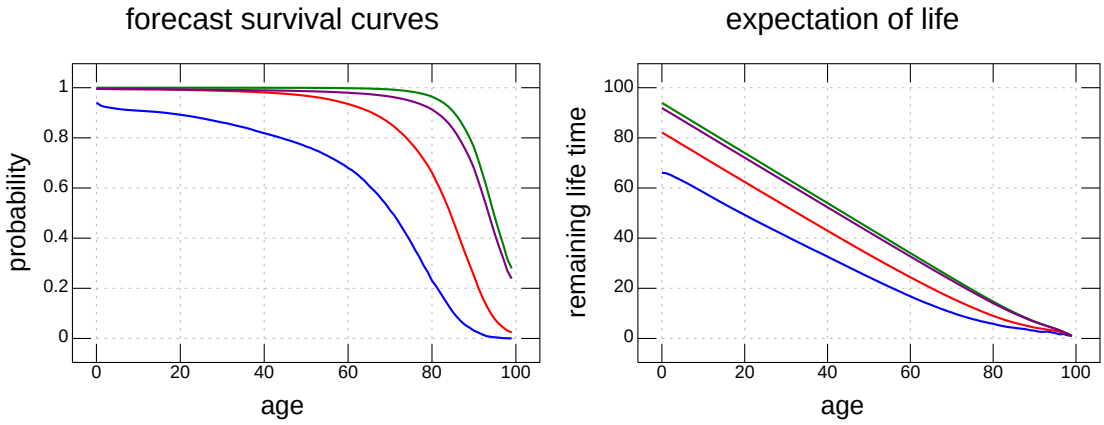

Figure 7: Actual and forecast survival probabilities and life expectancies

to age 100 at different ages. The bottom curve is computed from the 1921 life table. The next one up is from the 2000 table while the highest corresponds to the projected year 2100 table. Barely distinguishable from the 2100 table are the projected lifetime expectancies calculated for the 2000 cohort. Since the curves do not count lifetime beyond age 100, actual lifetimes are forecast to be on average, longer than those displayed, especially for the upper two curves where the probability of surviving beyond age 100 is appreciable.

It was noted in the §5 that mortality at age 70 and beyond improved at a faster rate than at the younger ages. This spells even higher forecast survival rates and life expectancies. Of course all these predictions are based on trends established over the last 80 years, and do not factor in trend reversals or stabilization.

When $\sigma > 0$ forecast survival curves will be less than under $\sigma = 0$. Hence if the current trend in the z–score progression remains but there is uncertainty about the trend, then forecast survival probabilities and expectations of life will be less.

References

Arnold, L. (1974). *Stochastic Differential Equations: Theory and Applications*. New York: John Wiley.

Bowers, J. N., H. Gerber, J. Hickman, D. Jones, and C. Nesbitt (1997). *Actuarial Mathematics*. Schaumburg, IL: Society of Actuaries.

De Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer.

De Jong, P. (1989). Smoothing and interpolation with the state-space

model. *Journal of the American Statistical Association 84*(408), 1085–1088.

De Jong, P. and S. Mazzi (2001). Modelling and smoothing unequally spaced sequence data. *Statistical Inference for Stochastic Processes 4*(1), 53–71.

De Jong, P. and J. R. Penzer (1998). Diagnosing shocks in time series. *Journal of the American Statistical Association 93*(442), 796–806.

De Jong, P. and L. Tickle (2006). Extending the Lee Carter model of mortality projection.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Heligman, L. and J. H. Pollard (1980). The age pattern of mortality. *Journal of the Institute of Actuaries 107*, 49–80.

Lee, R. D. and L. W. Carter (1992). Modelling and forecasting U.S. mortality (with discussion). *Journal of the American Statistical Association 87*(419), 659–675.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models* (2d ed.). New York: Chapman and Hall.

McNown, R. and A. Rogers (1989). Forecasting mortality: A parameterized time series approach. *Demography 26*(4), 645–660.

Wang, S. S. (2000). A class of distortion operators for pricing financial and insurance risks. *The Journal of Risk and Insurance 67*(2), 15–36.