

**DEPARTMENT OF
ACTUARIAL STUDIES
RESEARCH PAPER
SERIES**

**Extending the Lee-Carter methodology of
mortality projection**

by

Piet de Jong and Leonie Tickle

Queries to pdejong@efs.mq.edu.au

Research Paper No. 2004/07
September 2004

Division of Economic and Financial Studies
Macquarie University
Sydney NSW 2109 Australia

The Macquarie University *Actuarial Studies Research Papers* are written by members or affiliates of the Department of Actuarial Studies, Macquarie University. Although unrefereed, the papers are under the review and supervision of an editorial board.

Editorial Board:

Piet de Jong

Jim Farmer

Copies of the research papers are available from the World Wide Web at:

http://www.actuary.mq.edu.au/research_papers/index.html

Views expressed in this paper are those of the author(s) and not necessarily those of the Department of Actuarial Studies.

Extending the Lee–Carter methodology of mortality projection

Abstract

This article discusses the fitting and use of the Lee–Carter model for mortality forecasting. Shortcomings of the standard approach are discussed and a more flexible approach is suggested based on standard time series approaches to estimation and forecasting. The approach is applied to female mortality data.

1 Introduction

This article deals with the analysis and forecasting of mortality using observed mortality time series. A typical data set is displayed in Figure 1 relating to the log–mortality of Australian females from ages from 0 through to 100 and calendar years 1921 through to 2000. This data set has previously been analyzed by Booth, Maindonald, and Smith (2001) and Booth, Maindonald, and Smith (2002). As expected the mortality is high in the year of birth, rapidly falls off and then steadily increases. Although not very apparent from the plot, mortality at each age generally declines over time.

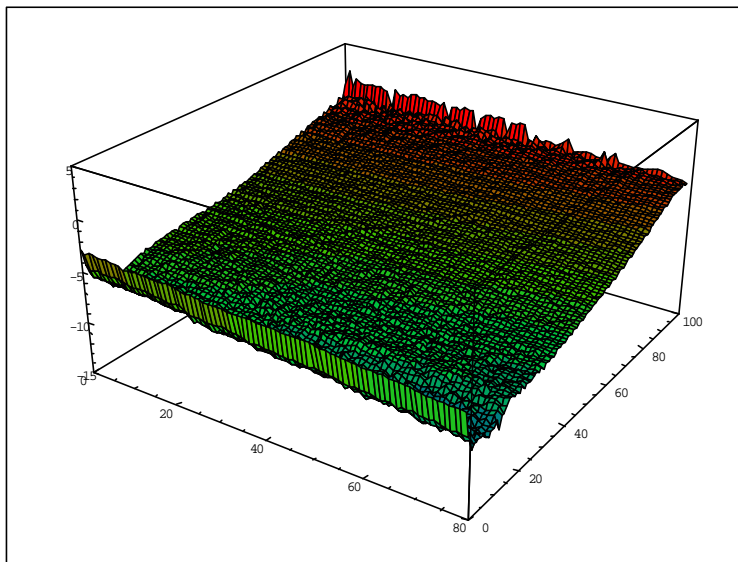


Figure 1: Log–mortality of Australian females, 1921 - 2000

The issue to be discussed in this paper is the forecasting of future mortality rates. One increasingly popular approach is based on the so–called “Lee–Carter” (LC) model (Lee and Carter 1992) which is both a model and an associated estimation method.

This paper aims to improve on the LC approach by strengthening both the modelling and estimation framework. Additional previous work related to the LC model not cited above includes Bozik and Bell (1987), Alho (1990), Bell and Monsell (1991), Bell (1992), Bell (1997), Lee and Miller (2001) and Currie, Durban, and Eilers (2002).

The next section applies LC and discusses areas of improvement. Suggested improvements and extensions occupy the rest of the paper. Issues with the singular value decomposition estimation approach are discussed in §3.1. An alternative least squares estimation method is developed and applied in §3.2. Another shortcoming is the over-parametrization of the raw form of the LC model implying fitted age related functions are likely to be too rough. Section 4 introduces a “smooth” version of the LC model, called LC(smooth), which retains the attractive features of LC as well as minimal parametrization and hence automatic smoothness. Section 5 discusses maximum likelihood estimation applicable even with more detailed time series specifications. Section 6 and 7 go on to discuss various extensions and specializations of the LC(smooth) model. The female mortality data of Figure 1 is used throughout the article to illustrate methods.

2 Overview of the Lee–Carter model and extensions

Let y_{it} be the logarithm of the central mortality rate at age $i = 0, 1, \dots, p$ and time $t = 1, \dots, n$. Put $y_t \equiv (y_{0t}, \dots, y_{pt})'$. Lee and Carter (1992) proposed the following model for studying and forecasting mortality – called the “Lee–Carter” (LC) model:

$$y_t = a + b\alpha_t + \epsilon_t, \quad t = 1, \dots, n. \quad (1)$$

Here a and b are $(p + 1)$ –vectors, of unknown parameters, ϵ_t is a vector of disturbance terms, and α_t is an unobserved time series process such as a random walk with drift. Here a measures the time invariant age effect on mortality, while b measures the age response to mortality trends. The parameters a and b , and the time series process α_t , are to be estimated.

Lee and Carter (1992) (see also Lee (2000)) propose the following estimate for the parameters in (1). Put $Y \equiv (y_1, \dots, y_n)$ implying Y is a $(p + 1) \times n$ matrix. Then a is estimated as the vector of row means of Y denoted \bar{y} with components equal to the average log–mortality over the period for each age. The singular value decomposition (svd) is then used on the mean corrected log–mortality rates: $Y - \bar{y}1' \simeq UDV'$ where 1 denotes a column vector of ones and \simeq indicates the retention of a small number of the largest singular values. Then $\hat{b} = U$ while $(\hat{\alpha}_1, \dots, \hat{\alpha}_n) = DV'$ and where hats indicate estimates.

The estimation approach is equivalent to first constructing the sample covariance matrix of observed log–mortality rates using age as variables and the time series as observations. Secondly, principal components from the sample covariance matrix are extracted and then the principal component time series corresponding to the largest principal component. This synthesized variable corresponds to $\hat{\alpha}_t$. This approach was pioneered in Bell and Monsell (1991).

Different authors have proposed using two or more components modelling mortality improvements. In this case α_t is a vector and b a matrix with as many columns as there are components in α_t .

A primary purpose of this paper is to discuss shortcomings and improvements to the

LC model. Improvements include the generalized specification

$$y_t = Xa + Xb\alpha_t + \epsilon_t . \quad (2)$$

Here X is a known “design” matrix with more rows than columns unless $X = I$ in which case (2) reduces to (1). As in (1), a and b are unknown while α_t is a time series process. We call this model the LC(smooth) model. The model (2) is developed and discussed in §4, after the evaluation of (1) in the next few subsections. In the models (1) or (2), smoothness across time results from imposing a time series structure on α_t .

2.1 Forecasting with the LC model

Forecasting mortality rates with (1) proceeds as follows. First an appropriate time series model is fitted to $\hat{\alpha}_t$. Components of α_t are then forecast into the future together with appropriate error bounds using standard time series methods. Multiplying the forecast components $\hat{\alpha}_{n+s}$ by the estimated age effects \hat{b} and adding \hat{a} gives the forecast age specific log-mortality at any point in the future. Thus the forecast log mortality rates out to year $n + m$ are

$$\hat{a}1' + \hat{b}(\hat{\alpha}_{n+1}, \dots, \hat{\alpha}_{n+m}) , \quad (3)$$

Cohort aged k in year $t = n$ is thus forecast to experience future log mortality displayed along subdiagonal $k+1$ of (3). Cohorts to be born at $t = n+k$ are predicted to experience mortality corresponding to superdiagonal $k - 1$ of (3). And the proportional change in mortality rate at the different ages out to year $n + m$ are forecast as approximately $\hat{b}(\hat{\alpha}_{n+m} - \hat{\alpha}_n)$.

2.2 Application to female mortality data

The top left panel of Figure 2 displays \bar{y} , and hence the base rates derived via the above method applied to the data in Figure 1. The top right panel displays the time series plots of the estimated α_t corresponding to the first two singular values. The major component is close to a straight line and suggests the virtual exponential decline in mortality over the period. The second AR(1) component peaks towards the middle of the period. This second component in this case is far less important than the first indicating its smaller significance as an explanation of historical trends in female mortality. In particular, the standard deviation associated with the first component is 6.07 times that of the second component, and the two components together account for over 99% of the total variance in the historical log-mortality.

The bottom left panel of Figure 2 graphs the age response profiles and hence second and third columns of the estimate of b constructed via the svd method. The plot lying above the x -axis and moving down to the right is the age response to the first time series component which reflects the general improvement in mortality over time. Thus the younger ages appear the most responsive to the improving trend in mortality. The downward bump around age 17 perhaps indicates the decreased response on account of increased traffic related mortality. The second graph roughly differentiates those in the population between the ages of about 13 through to 38 and the rest. Combining the information with that of the second component suggests that this fertile age group appears to have been subject to decreases in mortality over and above the general decrease experienced by all ages, in the period from about 1945 through to the late 70's.

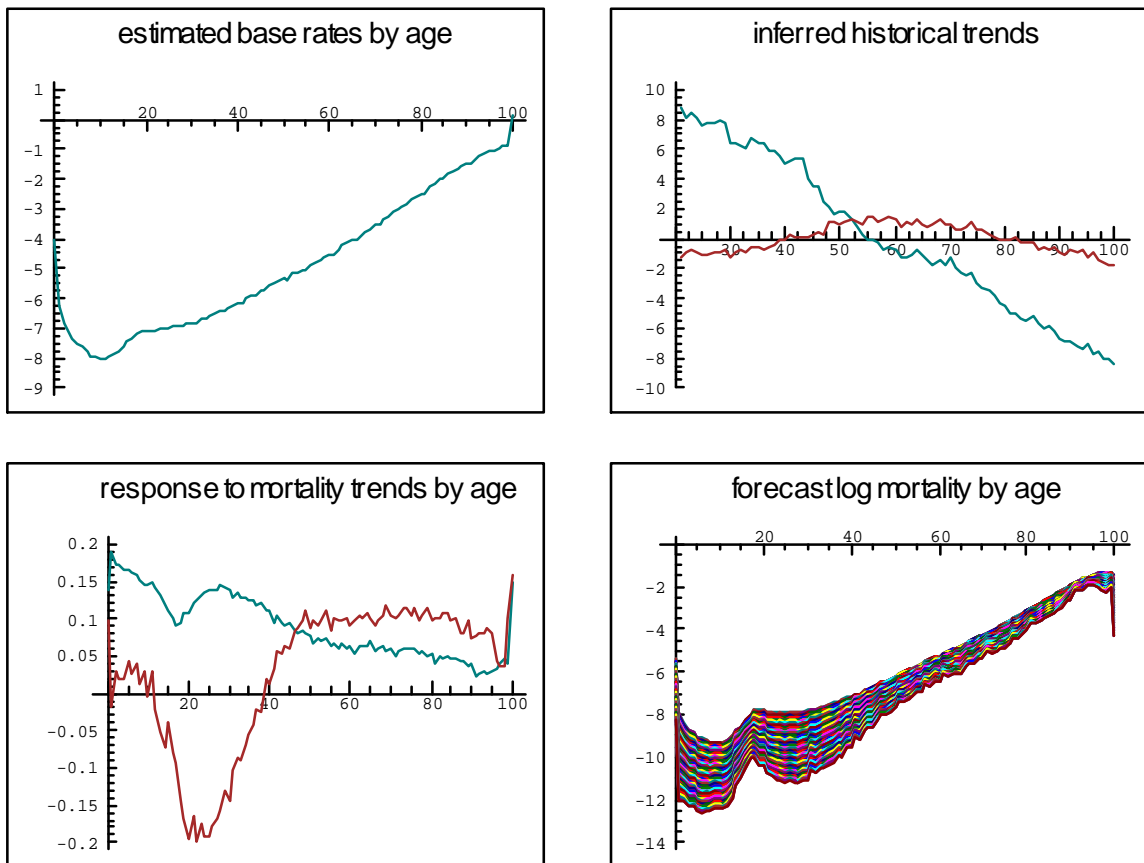


Figure 2: Results from two component analysis of log-mortality data

For the female mortality data, standard time series methods suggest that the first inferred time series component is adequately described by a random walk with drift of $\hat{\delta} = -0.218$ and noise standard deviation $\hat{\sigma}_1 = 0.370$. Since the average age effect corresponding to the first component is 0.090, this translates to an average improvement in mortality, averaging across all ages, of about $0.09 \times 0.218 = 1.96\%$ per annum. The second time series component appears well described by an AR(1) model with autoregressive coefficient $\hat{\phi} = 0.92$ and error standard deviation $\hat{\sigma}_2 = 0.26$.

3 Improving the LC methodology: estimation

This article suggests improvements to the LC model. Suggested improvements are the LC(smooth) model (2) and improving the estimation and forecasting basis. This section begins by examining potential shortcomings of the svd estimation technique usually applied in the context of the LC model (1) and secondly suggests a least squares improvement. This least squares improvement is a first step towards likelihood based estimation as outlined in §5.

3.1 Features of the svd estimation method

In the model (1), a_i is estimated as the average, over the period, of the log-mortality rate at age i : $\hat{a}_i \equiv \bar{y}_i$, $i = 0, \dots, p$ where \bar{y}_i is the mean of row i of Y . The purpose of this section is to demonstrate and discuss the further simple relations for the single component svd estimates of b_i and α_t , namely

$$\text{cor}(\hat{b}_i, s_{i\cdot}) \approx 1, \quad \text{cor}(\hat{\alpha}_t, \bar{y}_t) \approx 1, \quad (4)$$

where \bar{y}_t is the mean of column t of Y and $s_{i\cdot}$ is the standard deviation of row i of Y and cor denotes correlation.

Insight into the accuracy of (4) is gained from the two panels of Figure 3 which plot relevant quantities for the female mortality data. The first panel graphs \hat{b}_i against $s_{i\cdot}$ for $i = 0, \dots, 100$ while the second panel plots $\hat{\alpha}_t$ against \bar{y}_t for $t = 1, \dots, 80$. Clearly, there is a very strong linear relationship and hence the approximations in (4) are excellent. The conclusion derived from (4) is that for practical purposes the svd estimates reduce to simple means and standard deviations. Given this perspective, the svd estimates must be viewed with considerable skepticism as a basis for mortality projection and forecast error calculations.

An explanation of the approximations in (4) is as follows. Put $M \equiv Y - \bar{y}1 \simeq UDV'$ where UDV' is a one component singular value decomposition. Then $\hat{b} = U$ and $\hat{\alpha} \equiv (\hat{\alpha}_1, \dots, \hat{\alpha}_n)' = VD$ are eigenvectors of MM' and $M'M$ respectively, corresponding to the largest eigenvalue, D^2 . These eigenvectors maximize $b'MM'b$ and $\alpha M'M\alpha$ with respect to b and α , respectively, subject to appropriate normalization. Thus \hat{b} is the normalized linear combination of age specific mortality which has maximum variance moving across the calendar years. Since M has mean zero along each row, diagonal entries of MM' are proportional to the estimated variances $s_{i\cdot}^2$. Further, component \hat{b}_i tends to be large whenever $s_{i\cdot}$ is relatively large, leading to the first relation in (4). Thus the age-response estimates \hat{b}_i are essentially a measure of standard deviation in log-mortality for that age, across time.

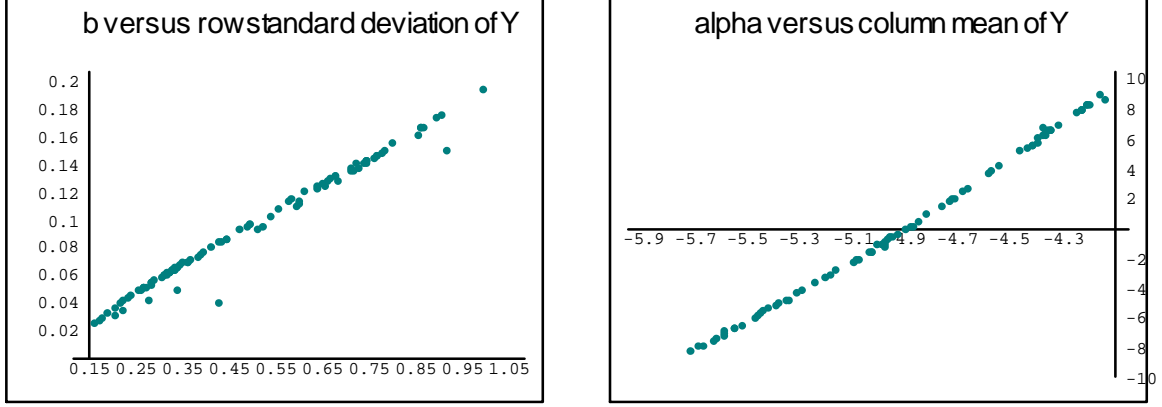


Figure 3: Plots of (s_i, \hat{b}_i) and $(\bar{y}_t, \hat{\alpha}_t)$ for female mortality data

For the second relation in (4), note that the columns of M are not mean corrected and if m is the vector of column means then

$$m \equiv (\bar{y}_1 - \bar{y}, \dots, \bar{y}_n - \bar{y})', \quad \alpha' M' M \alpha = \alpha' (M - 1m)' (M - 1m) \alpha + (p + 1)(\alpha' m)^2,$$

where \bar{y} is the mean of all the entries in Y . Maximizing the second expression with respect to the components of α yields the (unnormalized) eigenvector of $M' M$. Clearly the eigenvector will load heavily on all components t where $m_t \equiv \bar{y}_t - \bar{y}$ is large. This leads to the second approximation in (4). Thus $\hat{\alpha}_t$ is essentially the average log-mortality over all ages in calendar year t .

3.2 Least squares estimation of the LC model

One improvement to the “raw” implementation of the LC model as discussed in §2 is to integrate the estimation of the time series structure with the estimation of the age related parameters. This section displays a least squares approach. This sets the stage for generalized and improved estimation methods outlined in subsequent sections.

Consider the random walk with drift model

$$y_t = a + b\alpha_t + \epsilon_t, \quad \alpha_{t+1} = \delta + \alpha_t + \lambda\eta_t, \quad \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim (0, \sigma^2 I) \quad (5)$$

This model has two indeterminacies. First, multiplying b by a constant and dividing α_t by the same constant leaves the model unchanged. Thus without loss of generality assume $b'b = 1$. Second, replacing α_t by $c + \alpha_t$, $t = 1, \dots, n$ and simultaneously replacing a by $a - cb$ also leads to the same model and hence without loss of generality assume $\alpha_0 = 0$ or some other convenient value. Note that λ is a “signal-to-noise ratio” statistic.

From (5)

$$E(y_t | \alpha_0) = a + b(\alpha_0 + \delta t), \quad E(\Delta y_t) = \delta b. \quad (6)$$

Straightforward calculations show

$$\text{cov}(y_t, y_s | \alpha_0) = \begin{cases} \sigma^2 \{I + \lambda^2 t b b'\}, & t = s, \\ \sigma^2 \lambda^2 \min(t, s) b b', & t \neq s, \end{cases} \quad (7)$$

$$\text{cov}(\Delta y_t, \Delta y_s) = \begin{cases} \sigma^2(2I + \lambda^2 bb'), & t = s, \\ -\sigma^2 I, & t = s \pm 1, \\ 0, & |t - s| > 1. \end{cases} \quad (8)$$

Unbiased estimators of δb and $\text{cov}(\Delta y_t)$ are thus

$$\widehat{\delta b} = \frac{1}{n-1} \sum_{t=2}^n \Delta y_t = \frac{1}{n-1} (y_n - y_1), \quad \frac{1}{n-1} \sum_{t=2}^n (\Delta y_t - \widehat{\delta b}) (\Delta y_t - \widehat{\delta b})'. \quad (9)$$

Using the normalization $b'b = 1$ yields the estimates

$$\hat{\delta} = \frac{1}{n-1} \sqrt{(y_n - y_1)'(y_n - y_1)}, \quad \hat{b} = \frac{y_n - y_1}{\sqrt{(y_n - y_1)'(y_n - y_1)}}. \quad (10)$$

Furthermore $\bar{y} \approx a + b\{\alpha_0 + \delta(n+1)/2\}$ given \hat{b} , and hence if α_0 is normalized such that $\alpha_0 = -\hat{\delta}(n+1)/2$, then $\hat{a} = \bar{y}$ and $\hat{\alpha}_t = \hat{b}'(y_t - \bar{y})$ are reasonable estimates of a and α_t , respectively. The normalization on α_0 facilitates direct comparison to the estimates displayed in §2.

From (7) we find

$$\text{cov}(\widehat{\delta b}) = \sigma^2 \left\{ \frac{2}{(n-1)^2} I + \frac{\lambda^2}{n-1} bb' \right\} \approx \frac{\sigma^2 \lambda^2}{n-1} bb'. \quad (11)$$

Finally, note from (8) that since $b'b = 1$ we find that $\text{cov}(b'\Delta y_t) = \sigma^2(2 + \lambda^2)$ and $\text{cov}(b'\Delta y_t, b'\Delta y_{t-1}) = -\sigma^2$ suggesting the moment estimates

$$\hat{\sigma}^2 = \frac{-1}{n-1} \sum_{t=2}^n (\hat{b}'\Delta y_t - \hat{\delta})(\hat{b}'\Delta y_{t-1} - \hat{\delta}), \quad \hat{\lambda}^2 = - \left\{ 2 + \frac{\sum_{t=2}^n (\hat{b}'\Delta y_t - \hat{\delta})^2}{(n-1)\hat{\sigma}^2} \right\}. \quad (12)$$

These estimates are not necessarily positive, and, if negative, suggest the model is inappropriate.

3.3 Least squares estimates for female mortality data

Figure 4 illustrates the use of the least squares method applied to the female mortality data. The estimated base rates are, by construction, the same as those displayed in Figure 2. The estimated historical trend $\hat{\alpha}_t$ is displayed in the top right panel and this graph gives broadly the same conclusions as that trend estimated via the svd method. The response profile \hat{b} in the bottom left panel is smoother to that derived via the svd estimate. Also it is quite jagged arguing for appropriate smoothing. The final panel are forecast mortality rates for the next 100 years. These forecast rates also display roughness which perhaps would not be expected of actual rates. Finally $\hat{\delta} = -0.220$, $\hat{\sigma} = 0.155$ and $\hat{\lambda} = 1.911$. Thus the drift is quite similar to that derived with the svd method. Further, the variability associated with the random walk component $\hat{\lambda}\hat{\sigma} = 0.296$ and is almost twice that of error standard deviation associated with each observed log-mortality.

The difference in the least squares estimate of the random walk components and the estimate associated with the largest singular value derived via the svd method is marginal. This indicates, following from §3.1, the least squares estimate of the random walk is essentially a scaled version of the column means of Y . The least squares estimate of b in (10) appears ripe for improvement, relying substantially on just the first and last

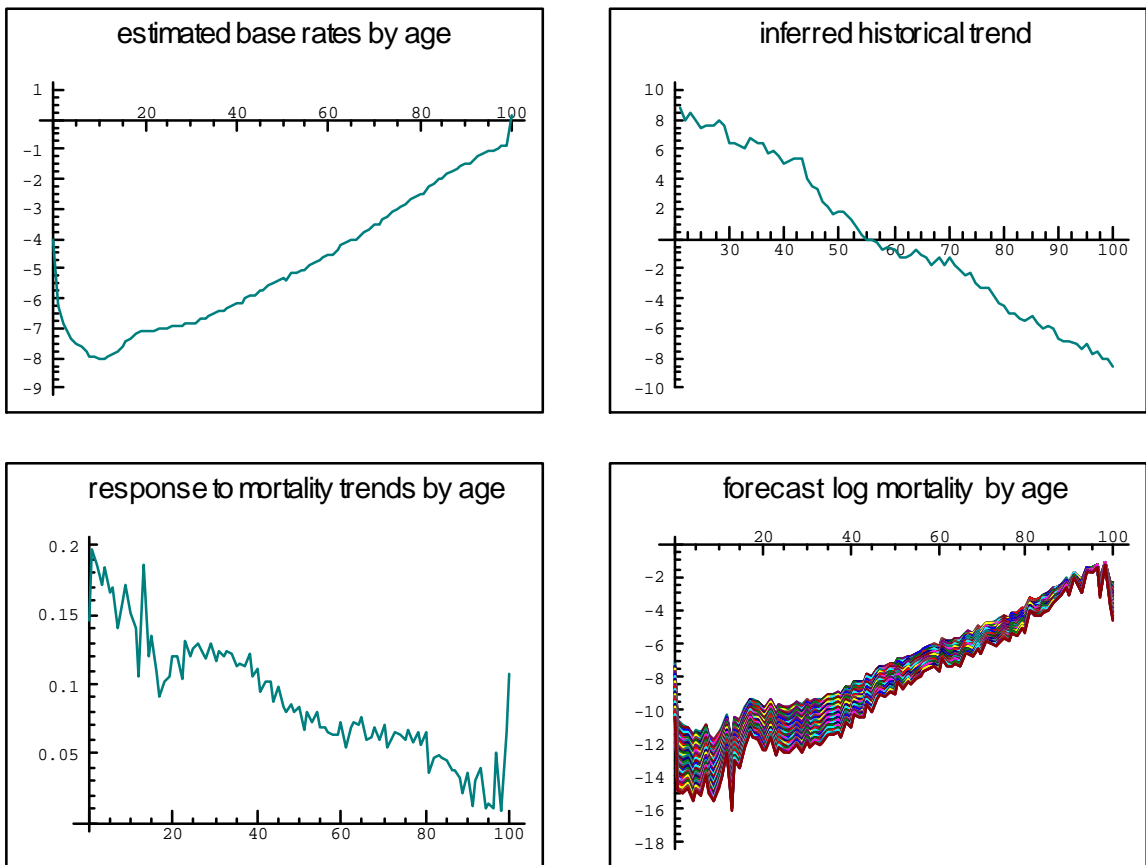


Figure 4: Result of least squares analysis for female mortality data

year's mortality experience. One suggestion is to regress y_t on $(1, t)$ leading to estimates of a and δb . However if allowance is made for the increasing in t covariance matrix of y_t as implied by the random walk model, then regression leads to the same estimates (10). Further, it appears worthwhile to iterate the estimates, improving efficiency. With each iteration, the estimate \hat{b} together with an estimate of λ is used to define the estimate of $\text{cov}\{\text{vec}(\Delta Y)\}$ as in (8) and covariances are used to improve estimates. A unified approach to this is discussed in §5.

4 The LC(smooth) model

This section discusses and applies a smoothed version of the LC model called the LC(smooth) model (2) where X is known. The LC(smooth) model builds in the expected smooth behavior of mortality by age. In practice X will have relatively few columns compared to rows. Thus a and b in (2) will have relatively few parameters as compared to a and b in the LC model (1). Thus the effect of X is to impose smoothness and robustness.

An example X is where entry (i, j) is $(i + 1)^{j-1}$, $i = 0, \dots, p$. Then X interpolates low order polynomials and the components of α_t impact mortality at each t in a smooth way across the ages. A variant is where X contains splines or b-splines (De Boor 1978), making for polynomial smoothness across age but allowing for knots at critical ages corresponding to break points in the momentum of age related mortality. In contrast to (1), the effects of the time series components in α_t across the ages are not functionally independent, but are smoothly constrained in terms of the variables making up the matrix X .

4.1 Least squares estimation of the LC(smooth) model

To estimate the LC(smooth) model (2), an initial least squares estimate can be obtained by transforming the model to a lower dimension based on X . The suggested transformation is based on $X = UDV'$, the svd of X , where it is assumed that V and D are square indicating the retention of only the nonzero singular values of X . Premultiplying the LC(smooth) model (2) by U' yields the reduced dimension LC model (1)

$$y_t^* = a^* + b^* \alpha_t + \epsilon_t^* , \quad (13)$$

where

$$y_t^* = U' y_t , \quad a^* \equiv U' X a = D V' a , \quad b^* \equiv U' X b = D V' b , \quad \epsilon_t^* = U' \epsilon_t ,$$

with $\text{cov}(\epsilon_t^*) = \sigma^2 I$ provided $\text{cov}(\epsilon_t) = \sigma^2 I$. The number of components in y_t^* is much less than y_t since X has few columns compared to rows. Thus the least squares method of §3.2 can be applied to y_t^* ensuring smoothness and robustness in the age direction. Estimates of a and b are arrived at by premultiplying the estimates of a^* and b^* by $(U'X)^{-1} = VD^{-1}$. Further, estimates of Xa and Xb are arrived at by premultiplying the estimates of a^* and b^* by U .

4.2 Application of the LC(smooth) to female mortality data

Figure 6 presents the results of fitting the LC(smooth) model (2) to the female mortality data of Figure 1 using the least squares method of §2. The matrix X used b-splines

graphed in Figure 5 and constructed as follows. Between knots log-mortality is assumed to follow a quadratic polynomial. Knots were placed at the following ages:

$$0.5, \quad 9.5, \quad 19.5, \quad 60.5, \quad 90.5, \quad 105.$$

At knots the piecewise quadratics join up in such a way that the derivative is continuous but may display “kinks” corresponding to discontinuities in derivatives. Knots correspond to apparent changes in the momentum of age related mortality. The values of the 8 splines are graphed in Figure 5. At each age there are at most 3 nonzero splines. Thus each row of X in (2) has at most 3 nonzero entries. The selected knots imply there are 8 age parameters associated with each time series component. This is a vast improvement on the $p + 1 = 101$ parameters, for each time series component, implicit in the analysis of §2.

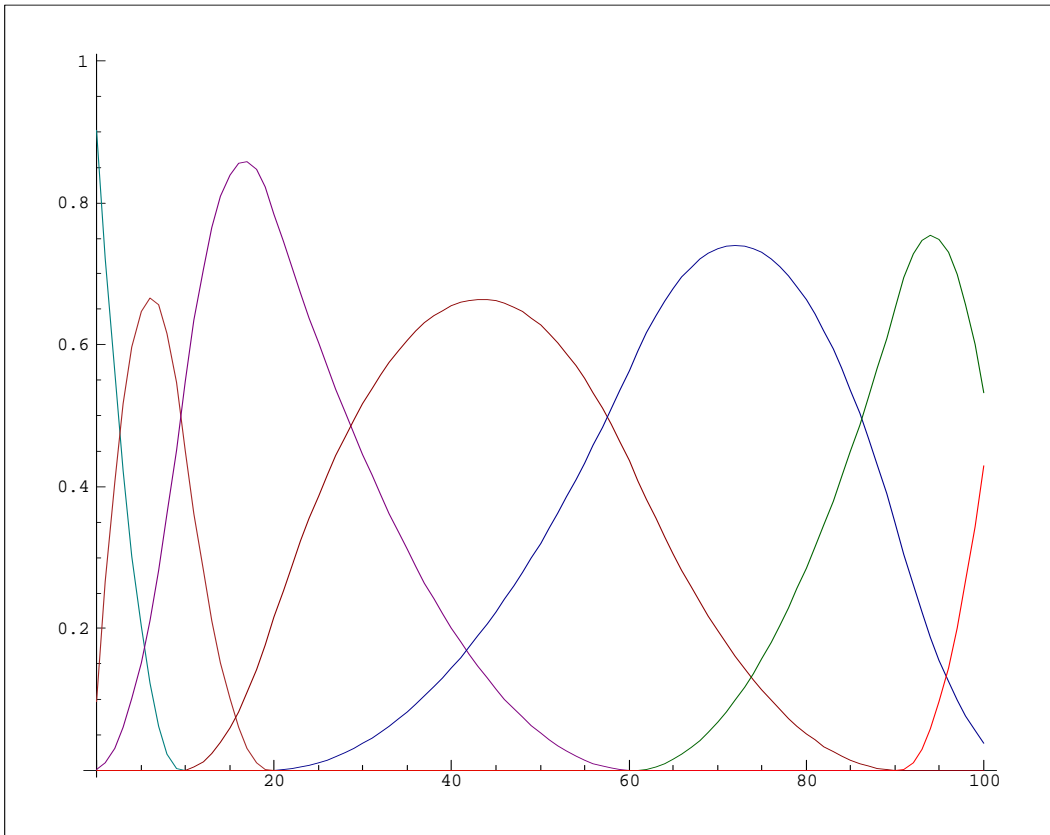


Figure 5: b-splines used for LC(smooth) applied to female mortality data

Figure 6 displays the results of the fit to the female mortality data. Both the base rates and the age response profile are smooth as a function of age. On the other hand the inferred historical trend is jagged. This is largely the result of the erratic behavior over time of the coefficients of the b-splines belonging to the extreme ages. This suggests the need to factor in standard errors associated with different ages.

4.3 Diagnostics

Insight into appropriateness of (2) is gained by considering $y_t^* = U'y_t$ and $N'y_t$ where the columns of N span the null space of the columns of X . If (2) holds, each of the

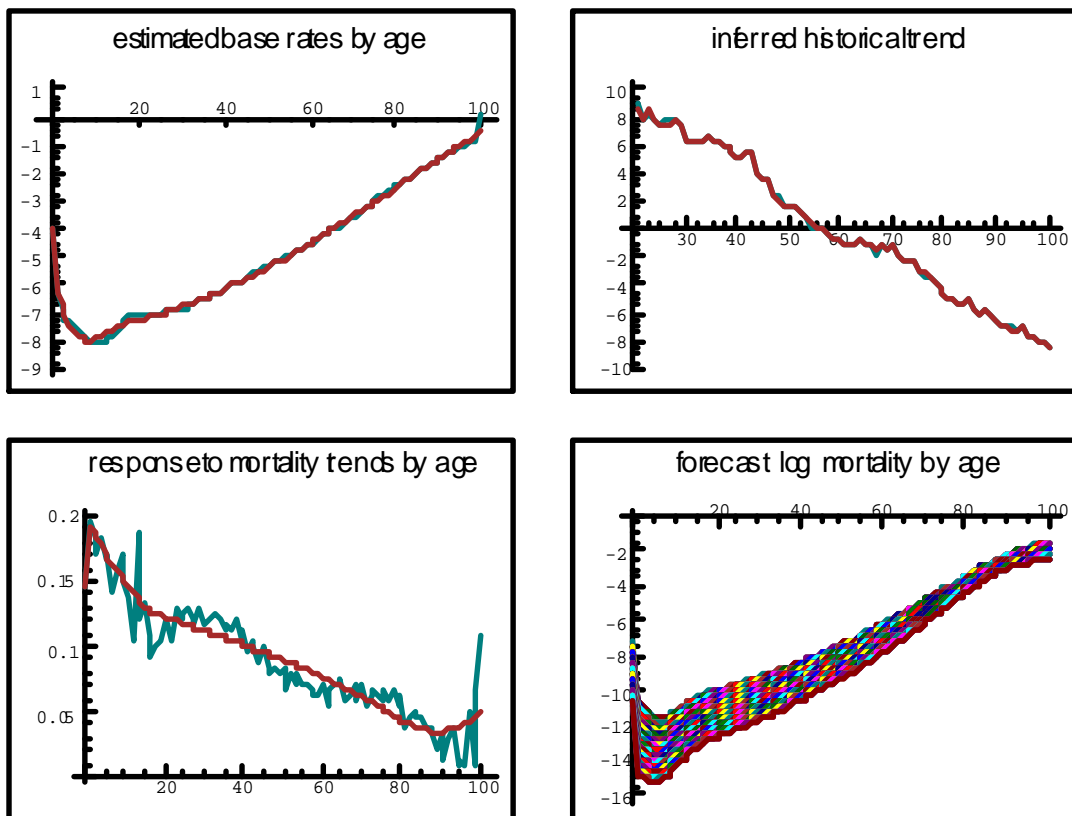


Figure 6: LC(smooth) least squares estimation results for female mortality

time series in y_t^* is a common random walk plus individual noise. The time series in $N'y_t$, on the other hand, provided $\text{cov}(\epsilon_t) = \sigma^2 I$ and $N'N = I$, are zero mean noise since $N'X = 0$. For example serial correlation in $N'y_t$ indicates not all the time dependence in log-mortality is picked up through the smooth age dependent behavior modelled with X .

The top left panel in Figure 7 plots the y_t^* for the female mortality data. The data is normalized so that $y_1^* = 0$. All seven series generally decline indicating mortality improvements at all age ranges. The least improvement occurs at the older ages. The top right panel indicates the autocorrelation function of the series in y_t^* . Five of the autocorrelation functions are consistent with that of an autoregressive process with a unit or near unit root and hence reasonably consistent with a common trend model that is a random walk. The final two autocorrelation functions, corresponding to the oldest ages, is not consistent with this specification. The partial autocorrelation functions are plotted in the bottom left panel of Figure 7 and also suggest AR(1) processes. An extension that replaces a common trend with separate trends is discussed in §7.

The bottom right panel indicates the average autocorrelation function of $N'y_t$ (middle line) as well plus and minus two standard deviations computed at each lag. The average autocorrelation suggests non-noise behavior for the series in $N'y_t$. In other words not all the predictability in y_t is picked up through the smooth age behavior modelled with the X matrix, suggesting more elaborate X matrix specifications.

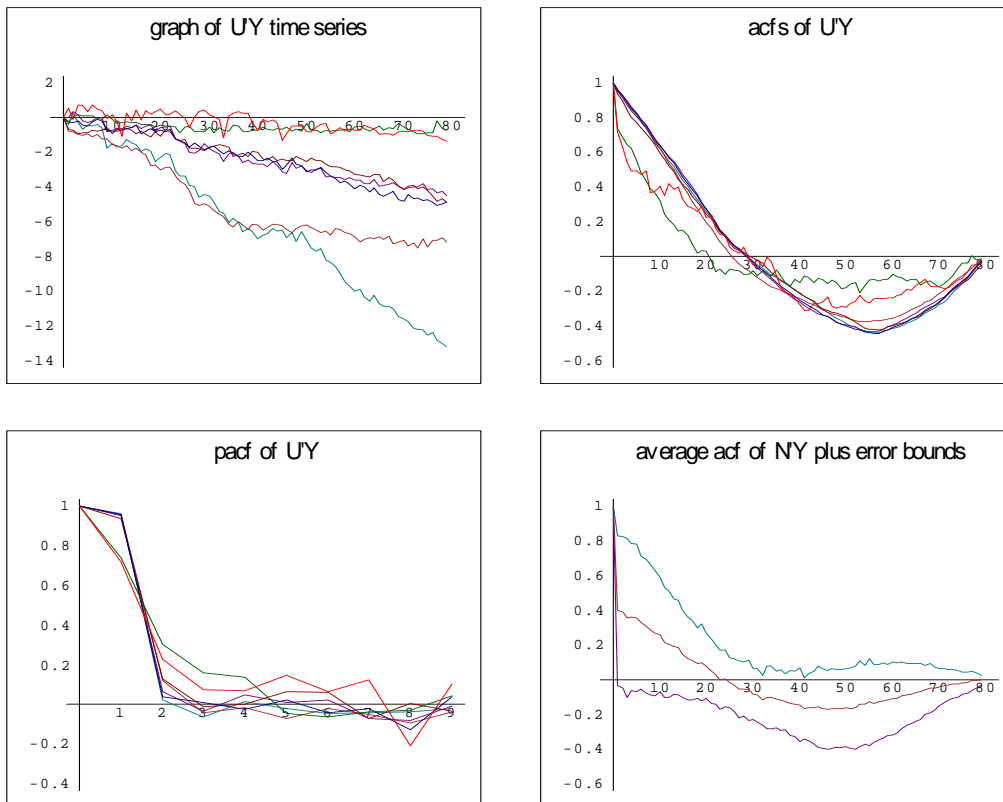


Figure 7: Time series properties of $y_t^* = U'y_t$ and $N'y_t$.

5 Maximum likelihood estimation of the LC(smooth) model using Kalman filtering

Maximum likelihood estimates of the LC(smooth) model (2) combined with an appropriate time series model for α_t can be derived using Kalman filtering and smoothing. To outline ideas, assume the time series model is a random walk with drift: $\alpha_{t+1} = \delta + \alpha_t + \lambda\eta_t$ and that disturbance terms are normally distributed.

1. Evaluate the normal based likelihood for different values of the hyperparameters b and λ using the Kalman filter and determine the maximum likelihood values \tilde{b} and $\tilde{\lambda}$.
2. Given \tilde{b} and $\tilde{\lambda}$ determine the maximum likelihood values of a , δ and σ , denoted \tilde{a} , $\tilde{\delta}$ and $\tilde{\sigma}$, again using the Kalman filter.
3. Given the maximum likelihood estimates of the parameters use the the Kalman and smoothing filters to determine

$$\tilde{\alpha}_t \equiv E(\alpha_t | y_1, \dots, y_n), \quad X(\tilde{a} + \tilde{b}\tilde{\alpha}_t) = E(Xa + Xb\alpha_t), \quad t = 1, \dots, n + s$$

and associated error variances. Here $s > 0$ indicates the the appropriate lead time for forecasting.

Kalman filtering can use either (2) or (13). Using (13) is faster since it operates in lower dimension. However (13) does involve loss of efficiency since the data $Y \equiv (y_1, \dots, y_n)$ is summarized into fewer observations $U'Y = (y_1^*, \dots, y_n^*)$. The numerical maximization in step 1 can be implemented using standard numerical routines. The examples below employ the simplex method (Nelder and Mead 1965) starting off from the least squares estimates discussed in §4.1.

The Kalman filtering strategy is applicable whenever the time series model for α_t is of the generic “state space” form. The random walk model $\alpha_{t+1} = \delta + \alpha_t + \lambda\eta_t$ is a particular instance. Other models include ARIMA or Box–Jenkins models (Box, Jenkins, and Reinsel 1994) or the structural models of Harvey (1989). The latter reference has an extensive discussion of state space models, Kalman filtering and likelihood maximization. Section 6 discusses more extensive time series model specifications amenable to Kalman filtering.

5.1 Application to female mortality data

Figure 8 reports the results using maximum likelihood estimation of the model (2) combined with a random walk with drift for α_t . The first panel compares the estimated base rates with those from a least square analysis of the model (2) originally shown in Figure 4. The difference in location results from a slightly different normalizations used for α_0 , in the present case $\alpha_0 = 0$. The shapes of the estimates are essentially the same. The top right panel indicates the estimate of the time series structure plus and minus two standard deviations, derived via the smoothing filter. The estimate is considerably smoother than the estimates derived via least squares and displayed previously. The bottom left panel displays the maximum likelihood estimate of Xb and compares it to the least squares estimate. The maximum likelihood estimate is the slightly more wavy

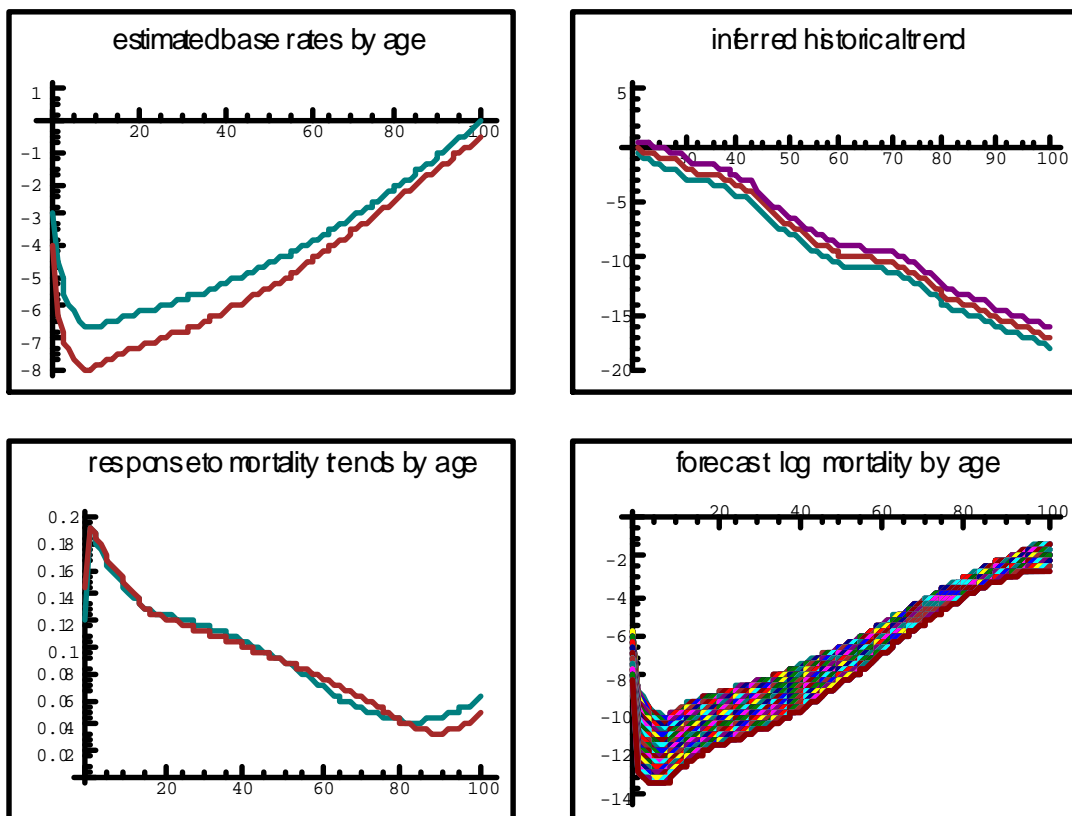


Figure 8: Maximum likelihood estimation results for female mortality data

Table 1: Comparison of estimates derived via different methods

	δ	σ	λ
Least squares	-0.220	0.155	1.911
Smooth Least squares	-0.218	0.148	1.983
mle using y_t^*	-0.216	0.359	0.661

line, but there is little difference between the two estimates. The bottom right panel displays the projected log-mortality for the next 100 years assuming drop

Table 5.1 compares estimates derived using the least squares, smooth least squares and mle methods. Each method estimates the drift δ to be around 0.22. Thus the actual expected improvement per annum is estimated to be $0.22 \times Xb$ where the estimate of Xb is plotted in the bottom left panel of Figure 8. At age 10 for example annual improvement is estimated to be about $0.22 \times 0.15 = 3.3\%$ per annum while the improvement at the older ages, say around 70, is estimated to be around $0.22 \times 0.05 = 1.1\%$ per annum.

6 Extensions and specializations of the LC(smooth) model

The LC(smooth) model (2) reduces to the LC model if $X = I$. Further specializations and extensions are discussed in the next few subsections.

6.1 Evolving mortality trend

In the standard setup δ , the approximate per period percentage change in mortality, is fixed, independent of t . More generally mortality trends may vary over time leading to

$$\begin{pmatrix} \alpha_{t+1} \\ \delta_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_t \\ \delta_t \end{pmatrix} + \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \quad (14)$$

The case $\lambda_2 = 0$ corresponds to a constant mortality improvement. The general model is appropriate in situations where the rate of decline in log-mortality is not constant. The effect of “slope” noise η_{2t} and hence $\lambda_2 > 0$ is to ensure the most recent data is of more importance in determining mortality trends. An extensive discussion of models of the form (14) is contained in Harvey (1989, p.47).

6.2 More detailed time series modelling

An example is adding an AR(1) component to the random walk:

$$\alpha_{t+1} = \begin{pmatrix} \delta \\ \mu \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & \phi \end{pmatrix} \alpha_t + \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}, \quad (15)$$

In this case b is a matrix with two columns indicating the log mortality response profiles to the random walk and AR(1) components. More general models that can also be considered include ARIMA or Box-Jenkins models (Box, Jenkins, and Reinsel 1994) or the extended “structural” models of Harvey (1989).

6.3 Differential measurement error variances

Variability in log-mortality rates are expected to be different at the different ages. This can be allowed for in the calculation by replacing ϵ_t in (2) by $G\epsilon_t$ where $G \neq I$. In the estimation, high variability ages will weigh less heavily in determining trends. This is valuable in avoiding tedious other adjustments of α_t .

6.4 Constraining fitted rates.

The fitted rates may be constrained so that, for example the sum of the observed log rates is equal to the sum of the fitted log rates at each t . For example define $M \equiv I - n^{-1}\iota\iota'$ where $\iota = (1, \dots, 1)'$. Then $\iota'M = 0$ and if

$$y_t = Xa + Xb\alpha_t + M\epsilon_t, \quad (16)$$

then $\iota'y_t = \iota'(Xa + Xb\alpha_t)$. A fit based on (16) will lead to fitted rates $\tilde{y}_t \equiv X\tilde{a} + X\tilde{b}\tilde{\alpha}_t$ such that $\iota'y_t = \iota'\tilde{y}_t$ and hence the fitted ‘‘jump off’’ rates $\tilde{y}_n \equiv X\tilde{a} + X\tilde{b}\tilde{\alpha}_n$ for forecasting are, on average, equal to the average of the latest actual rates y_n . Smaller groups of fitted rates may be similarly constrained. This is achieved by replacing M in (16) by a block diagonal matrix with smaller versions of M on each block. These type of constraints can also be imposed on the transformed system: $y_t^* = a^* + b^* + M\epsilon_t^*$ in which case sums of fitted rates in the transformed system coincide with sums of components of y_t^* .

7 Separate trends

The LC(smooth) model states that a single common random walk drives the behavior over time of all the age specific mortality rates. This can be relaxed so that a number of random walks operate allowing, for example, the dynamic behavior of the earlier ages to differ from those in the latter years. This leads to the model

$$y_t = Xa + X\alpha_t + \epsilon_t, \quad \alpha_{t+1} = \delta + \alpha_t + H\eta_t \quad (17)$$

where α_t is a vector time series process. In effect the scaled versions of a single random walk $b\alpha_t$ in LC(smooth) (2) is replaced by a vector of random walks α_t . The model (17) specializes to the LC(smooth) model (2) by constraining δ and H in (17) to be scalar multiples of a common vector.

The equations in (17) correspond to a model previously considered in De Jong and Boyle (1983) who applied the model to the mortality experience of a relatively small pension fund. In their setting, there was considerable variation in exposure at the different ages and points of time and these differences were allowed for in their model through heteroskedastic error terms ϵ_t . With large sample sizes, implicit in for example the female mortality data, exposures are uniformly large and death probabilities are small except for the extreme old ages suggesting the $\text{cov}(\epsilon_t) = \sigma^2 I$ assumption is reasonably appropriate.

Multiplying the first equation in (17) by U' , where $X = UDV'$ is the svd of X , and rearrangement yields the lower order system

$$y_t^* = a^* + \alpha_t^* + \epsilon_t^*, \quad \alpha_{t+1}^* = \delta^* + \alpha_t^* + H^*\eta_t, \quad (18)$$

where $a^* = DV'a$, $\alpha_t^* \equiv DV'\alpha_t$, $\epsilon_t^* = U'\epsilon_t$, $\delta^* = DV'\delta$ and $H^* = DV'H$. Letting each random walk component have its own error term and assuming $H^* = \text{diag}(\lambda^*)$, a diagonal matrix, leads to

$$\alpha_{t+1}^* = \delta^* + \alpha_t^* + \text{diag}(\lambda^*)\eta_t ,$$

where η_t is a vector of disturbance terms. Notice that $Ua^* = Xa$ and $U\alpha_t^* = X\alpha_t$.

The system can be estimated using maximum likelihood and the Kalman filter. Again, basing estimation on (18) as opposed to (17) will be convenient numerically but involve loss of efficiency since the data $Y \equiv (y_1, \dots, y_n)$ is summarized into fewer observations $Y^* \equiv U'Y$. Unknown hyperparameters of (18) include σ^2 and the entries of H^* . Given the mle of λ^* , the mle's of a^* , δ^* and σ^2 follow via closed formulas. In turn, since $Xa = Ua^*$ and $X\alpha_t = U\alpha_t^*$, the fitted values of y_t are determined directly from the estimates. Forecast future log mortality rates at time $t = n + m$ are

$$U(\tilde{a}^* + \tilde{\alpha}_n^* + m\tilde{\delta}^*) \approx UU'y_n + mU\tilde{\delta}^* \approx y_n + mU\tilde{\delta}^* .$$

In the right hand side, the latest observed rates y_n are used as “jump off” rates for the random walk projections rather than the latest smoothed rates $U(\tilde{a}^* + \tilde{\alpha}_n^*)$.

References

- Alho, J. (1990). Stochastic methods in population forecasting. *International Journal of forecasting* 6, 521–530.
- Bell, W. (1992). Arima and principal component models in forecasting age-specific fertility. In N. Keilman and H. Crujisen (Eds.), *National population forecasting in industrialized countries*, Amsterdam, pp. 159–177. Swets and Zeitlinger.
- Bell, W. (1997). Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. *Journal of Official Statistics* 13, 279–303.
- Bell, W. and B. Monsell (1991). Using principal components in time series modelling and forecasting age-specific mortality rates. In *Proceedings of the American statistical association, Social statistics section*, pp. 154–159.
- Booth, H., J. Maindonald, and L. Smith (2001). Age-time interactions in mortality projections: applying Lee-Carter to Australia.
- Booth, H., J. Maindonald, and L. Smith (2002). Applying Lee-Carter under conditions of mortality decline. *Population Studies* 56, 325–336.
- Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (1994). *Time Series Analysis: Forecasting and Control* (3d ed.). Englewood Cliffs, New Jersey: Prentice-Hall.
- Bozik, J. and W. Bell (1987). Forecasting age-specific fertility using principal components. In *Proceedings of the American Statistical Association, Social Statistics Section*, pp. 396–401.
- Currie, I., M. Durban, and P. Eilers (2002). Using p-splines to smooth two-dimensional poisson data. In *Proceedings of the 17th International Workshop on Statistical Modelling, Crete*, pp. 127–134.
- De Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer.

- De Jong, P. and P. Boyle (1983). Monitoring mortality: a state-space approach. *Journal of Econometrics* 23, 131–146.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
- Lee, R. D. (2000). The Lee–Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal* 4(1), 80–93.
- 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.
- Lee, R. D. and T. Miller (2001). Evaluating the performance of the Lee–Carter method for forecasting mortality. *Demography* 38, 537–49.
- Nelder, J. and R. Mead (1965). A simplex method for function minimization. *The Computer Journal* 7, 308–313.