Class: SY BSc

Subject : Statistical Risk Modelling 1

Chapter: Unit 4 Chapter 1

Chapter Name: Mortality Projection

Today's Agenda

1. Introduction

1. Methods of Mortality Projection

1. Methods based on Expectation

1. Methods based on Extrapolation
    1. Lee-Carter Model
    2. Age Period Cohort Model
    3. P-Splines

1. Methods based on Explanation

1. Sources of Errors in Mortality Projection

Covid.

U4 Ch1

mortality projection

F & U. File and use.

U & F Use and file.

mortality project

0 2035

43 years

2068

2025

43

22

$\boxed{43}$

$\boxed{65}$

100,000

premium

claims

Force of mort/CRM.

$\mu / m_{22,43}$.

m_x

μ_x

q_x

p_x.

methods based on expectation
extrapolation
explanation

Avg life expectn. age.

time ↓
age ↑

age

24 - 30

2025
22

2040
22

Pensions · ↑ survival Ins. co. ↑ liab

overestⁿ in mortality

underestⁿ.

50

insurance.

prudent

better.

underestⁿ

more people will

20 5

30 15

↑

30

10 20

40

overestⁿ.

mort ↑ liab ↑.

underestimⁿ

worse

~~inflation~~

0·0005·

0·0003

pensions

underestimⁿ mort

more no. will live

estimated higher liab

better.

overestimⁿ mort

less no will live

estimated lower liab

worse

1 Introduction

Why do we project future mortality?

- The projection of future mortality rates is vital for many actuarial purposes, most obviously life insurance and pensions.
- If its **estimates of future mortality are too** low, a life **insurance company may run into financial difficulties**. On the other hand, a company offering **pensions and annuities could become uncompetitive**
- If **estimates of future mortality are too high**, on the other hand, the opposite problems may arise: the financial commitments of a **pension scheme may outweigh its resources**, but a **life insurance company may offer uncompetitive rates** and lose business.
- Governments, too, need accurate forecasts of future mortality rates. The future population of a country, especially the older age population, depends very much on future mortality trends

2 Methods of Mortality Projection

Three methods used to project future mortality are:

1. Methods based on expectation
2. Methods based on extrapolation
3. Methods based on explanation

3

Methods based on Expectation

- In the past, mortality projections generally relied on an expectation that previous trends in mortality improvements would continue, accompanied by consultation with demographers and other experts.

- For much of the twentieth century this proved reasonably satisfactory.
- However, more recently these methods have generally underestimated improvements in mortality, particularly in the period 1990-2010.

- An alternative approach along these lines is to suppose that, by some future date, mortality will have reached some target level, and to examine different ways in which the schedule of age-specific mortality rates might move from their present pattern in order to achieve that target.

120,000

30

$\boxed{2025}$

60

2055

3 Methods based on Expectation

Official statistical agencies have traditionally based their mortality projections on simple expectations (for example a continued exponential decline in age-specific mortality rates).

The parameters of the future evolution of mortality have been set either by fitting deterministic functions to recent mortality trends, or by inviting experts to indicate how far and fast they anticipate mortality to fall, and to set ‘targets’.

3 Methods based on Expectation

One approach involves the use of reduction factors, $R_{x,t}$, which measure the proportion by which the mortality rate at age x , q_x , is expected to be reduced by future year t .

$$\alpha_x + (1 - \alpha_x)(1 - f_{n,x})^{t/n}$$

We can write:

$$\boldsymbol{R_{x,t} = \alpha_x + (1 - \alpha_x)\,(1 - f_{n,x})^{t/n}}$$

where α_x is the ultimate reduction factor, and $f_{n,x}$ represents the proportion of the total decline expected to occur in n years. Expert opinion is used to set the targets α_x and $f_{n,x}$.

The mortality rate at age x in future year t, $\boldsymbol{q_{x,t}}$, would then be expected to be equal to:

$$\boldsymbol{q_{x,t} = R_{x,t} \cdot q_x}$$

$$q_{x,t} = q_x \cdot R_{x,t}$$

x — age

$t \rightarrow$ time

3 Methods based on Expectation

- Methods based on expectation have been widely used in the past and have the **advantage of being straightforward and easy to implement**. However, in **recent decades they have tended to underestimate improvements, especially for male mortality.** One reason for this is that progress on reducing male mortality was slow during the 1950s, 1960s and 1970s because of lifestyle factors such as cigarette smoking.

- There are **theoretical problems, too, with targeting**. The setting of the 'target' is a forecast, which implies that the method is circular; and setting a target leads to an underestimation of the true level of uncertainty around the forecast.

- An **alternative approach to generating expectations is to ask the population as a whole about its health status and general well-being**. Self-reported health status has been found to be a good way of identifying 'healthy' and 'unhealthy' individuals, but it is not clear that it can provide useful information on future longevity at the population level.

Based on expectation

- Example-

Age	α_x	$f_{20,x}$
60	0.13	0.55
80	0.478	0.446

- In year 1992, q_x for age 60 = 0.005914 & age 80 = 0.0075464.
- Calculate q_{60} in 2002 and q_{80} in 1997

$q_{60,10} = 0.004220$ $\quad q_{80,5} = 0.007005$

$t = 10$, $n = 20$

$f_{20,x}$

① R_{60}

20 years

α_{60} $\quad f_{20,60}$

α_{80} $\quad f_{20,80}$

1992

$q_{80} = 1997$

$\left[\alpha_x + (1-\alpha_x)(1-f_{n,x})^{t/n}\right]$ 0.005914

$q_{60} = 2002$ $\quad q_{80} = 1997$

$q_{60,1}$

$q_{60,2}$

$q_{60,3}$ 60 $\xrightarrow{\text{20 years}}$ 55%

1 year 2 years 3

$q_{60,4}$ 70 $\xrightarrow{\text{20 year}}$

30 $\xrightarrow{\text{20 years}}$

$f_{20,60}$

$f_{20,70}$

$f_{20,30}$ t

4 Methods based on Extrapolation

α_x, $\beta_{n,x}$ → estimate

Stochastic models

- Deterministic approaches based on expectation have been largely superseded, other than for short-term forecasting. More advanced approaches use stochastic forecasting models.

x,

- In this section we describe some commonly used models, but we start with a discussion of the factors that apply in forecasting mortality.

x, t

period

c

t

- Can be classified using no. of factors taken in forecasting

Model type	Factor	Notation
One factor	Age	m_x
Two factors	Age-Period or Age-Cohort	$m_{x,t}$ or $m_{x,c}$
Two factors	Age-Period-Cohort	$m_{x,t,c}$

- In 2 factor models, it is generally harder to use age cohort compared to age period
- Data requirement is very heavy for age cohort models and also, if available, it may not be sufficient

4 Methods based on Extrapolation

Age, period and cohort factors

- In general, when forecasting mortality, the problem is to produce estimates of $m_{x,t}$, the central rate of mortality at age x at time t , for some future time period, based on data for $m_{x,t}$ over some past time period.
- The $m_{x,t}$, are defined on the basis of two factors: age x and time (period) t .
- If we define t to be the projection year and x to be the age reached during that projection year, then, for example, $m_{70,2045}$ is the expected mortality rate of those people who reach the age of 70 during the year 2045.

2025· 90

$m_{70,\,2045}$ 70· → 2045

4 **Methods based on Extrapolation**

Age, period and cohort factors

- Age and period can be combined to produce a third factor, the cohort, defined, say, on the basis of date of birth. Because a person aged x at time t will have been born at time t - x, age, period and cohort are not independent.

- Forecasting models can be classified according to the number of factors taken into account in the forecasting processes, as follows:
  One-factor models - Age (m_x)
  Two-factor models - Age, period ($m_{x,t}$) OR Age, cohort ($m_{x,c}$)
  Three-factor models - Age, period, cohort ($m_{x,t,c}$)

- In two-factor models, it has been usual to work with age and period. It is possible to work with age and cohort, but the cohort approach makes heavy data demands and there is the largely insoluble problem that recent cohorts have their histories truncated by the present day.

4 Methods based on Extrapolation

Age, period and cohort factors

- In general, most research has found that cohort effects are smaller than period effects, though cohort effects are non-negligible.
- Three-factor models also have the logical problem that each factor is linearly dependent on the other two. Various approaches have been developed to overcome this problem, though none are entirely satisfactory because the problem is a logical one not an empirical one.

4.1 The Lee-Carter Model

Lee Carter model

x, t : age-period

One of the most widely used models is that developed by Lee and Carter in the early 1990s. The Lee-Carter model has two factors, age and period, and may be written as follows:

Lee-Carter model

$$\boldsymbol{\log_e m_{x,t} = a_x + b_x k_t + \varepsilon_{x,t}}$$

$\boxed{\ln m_{x,t}}$ $\quad m_{x,t} = \exp(a_x + b_x k_t + \varepsilon_x)$

where: $m_{x,t}$ is the central mortality rate at age x in year t

a_x describes the general shape of mortality at age x (more exactly it is the mean of the time-averaged logarithms of the central mortality rate at age x)

b_x measures the change in the rates in response to an underlying time trend in the level of mortality k

k_t reflects the effect of the time trend on mortality at time t , and

$\varepsilon_{x,t}$ are independently distributed random variables with means of zero and some variance to be estimated.

Constraints imposed in the Lee-Carter model

The usual constraints are that $\sum_x \boldsymbol{b_x} = \mathbf{1}$ and $\sum_t \boldsymbol{k_t} = \mathbf{0}$

a_x = measured as the avg $\ln m_{x,t}$

$$\frac{1}{n}\sum_{t=1}^{n} \ln m_{x,t}$$

$$= \frac{1}{n}\sum_{t=1}^{n}\left[a_x + b_x k_t\right]$$

$$= \frac{1}{n}\left\{\sum_{t=1}^{n} a_x + b_x \sum_{t=1}^{n} k_t\right\}.$$

$$= \frac{1}{n}\left[n a_x + b_x \cdot \times 0\right]$$

$$a_x = \frac{1}{n}\sum_{t=1}^{n} \ln m_{x,t}$$

4.1 Estimation of the Lee-Carter Model

There are several approaches to estimating the Lee-Carter model:

- The original approach of Lee and Carter first estimated the a_x as the time-averaged logarithms of mortality at each age x. They then used singular value decomposition of the matrix of centred age profiles of mortality log_e $m_{x,t}$ - $\hat{a}_x$ to estimate the b_x and k_t . (Singular value decomposition is a way of decomposing a matrix into three component matrices, two of which are orthogonal and one diagonal.)

- Alternatively, the LC model can be fitted using the method of maximum likelihood subject to constraints (Macdonald et al., 2018). This estimation approach can be carried out using the gnm package in R. However, the estimated parameters from the gnm package do not satisfy the constraints that $\sum_x \boldsymbol{b_x}$ **= 1** and $\sum_t \boldsymbol{k_t}$ **= 0.** But, a simple adjustment of the estimates produced by the gnm function will recover estimates of a_x , b_x and k_t which do satisfy these constraints.

gnm

4.1 Forecasting with the Lee-Carter Model

The Lee-Carter model has three sets of parameters, a_x , b_x and k_t. Two of these relate to the age pattern of mortality, whereas the third, k_t, measures how mortality evolves over time. Forecasting with the model in practice involves forecasting the kt while holding the ax and bx constant.

The random walk model may be written as:

$$k_t - k_{t-1} = \Delta\, k_t = \mu + \varepsilon$$

where μ measures the average change in the kt and εt are independent normally distributed error terms with variance σ^2.

Suppose that t_0 is the latest year for which we have data. Having estimated μ using data for t < t0, forecasting can be achieved for the first future period, as:

$$\hat{k}_{t_0+1} = k_{t_0} + \hat{\mu}$$

a_x b_x k_t ε

4.1 Forecasting with the Lee-Carter Model

and in general, for l years ahead:

$$\hat{k}_{t_0+l} = k_{t_0} + l\hat{\mu}$$

Predicted future mortality rates in year t_0 + l are then obtained as:

$$\boxed{\boldsymbol{\log_e \hat{m}_{x,t_0+l} = \hat{a}_x + \hat{b}_x\, \hat{k}_{t_0+l}}}$$

$t_0 + l$ $\hat{k}_{t_0+l}$ l t $\hat{k}_t$

1. Lee Carter model

- Advantages
- Once parameters estimated, forecasting is straightforward
- Extent of error/degree of uncertainty in parameter can be estimated
- Has varied applications ex – for smoothing age patterns of mortality

- Disadvantages
- Parameters a_x and b_x are constants, and future estimates are heavily dependent on them
- Estimates may be distorted due to any roughness observed in past data (ex. Past events that make mortality uncertain)
- Forecasts become increasingly rough over time (due to error terms)
- It does not include a cohort term where evidence suggests that cohort based improvements are also non negligible

4.1 Extensions to The Lee-Carter Model

There is evidence that most of these modifications and alternatives to the original Lee-Carter model give lower forecast errors than the original in respect of age-specific mortality rates, though there is little difference when the expectation of life is considered.

born

cohort

The age-period-cohort model

An age-period-cohort extension of the Lee-Carter model may be written:

$$\log_e m_{x,t} = a_x + b_x^1 k_t + b_x^2 h_{t-x} + \varepsilon_{x,t}$$

where h_{t-x} is the overall level of mortality for persons born in year t-x.

In this case h_{t-x} can be estimated from past data and the forecasting achieved using time series methods similar to those described for the kt.

2045

age – x

period – t

cohort c = t-x.

4.1 Splines

A spline is a polynomial of a specified degree defined on a piecewise basis. The pieces join at knots, where certain continuity conditions are fulfilled to ensure smoothness. Earlier the splines were fitted to age-dependent mortality rates, so the knots were specified in terms of ages. Typically, the polynomials used in splines in mortality forecasting are of degree 3 (i.e. cubic).

To construct the model, we choose the number of knots (and hence the number of splines to use), and the degree of the polynomials in each spline. We can then use the splines in a regression model, such as the Gompertz model. To illustrate, the Gompertz model can be written as:

$$\log_e[E(D_x)] = \log_e E_x^c + \alpha + \beta x,$$

where E(Dx) is the expected deaths at age x, E_x^c is the central exposed to risk at age x, and α and β are parameters to be estimated.

If we replace the term α + βx by a smooth function defined using splines, we have:

$$\log_e[E(D_x)] = \log_e E_x^c + \sum_{j=1}^{s} \theta_j B_j(x),$$

where B_j(x) are the set of splines, and θ_j are the set of parameters to be estimated.

4.1 p-Splines

- Spline models will adhere more closely to past data if the number of knots is large and the degree of the polynomial in the spline function is high. For forecasting, we ideally want a model which takes account of important trends in the past data but is not influenced by short-term “one-off” variations.
- This is because it is likely that the short-term variations in the past will not be repeated in the future, so that taking account of them may distort the model in a way which is unhelpful for forecasting.
- On the other hand, we do want the model to capture trends in the past data which are likely to be continued into the future.

- Models which adhere too closely to the past data tend to be “rough” in the sense that the coefficients for adjacent years do not follow a smooth sequence.

- The method of p-splines attempts to find the optimal model by introducing a penalty for models which have excessive “roughness”.

4.1 p-Splines

The method may be implemented as follows

- Specify the knot spacing and degree of the polynomials in each spline.
- Define a **roughness penalty**, $P(\theta)$, which increases with the variability of adjacent coefficients. This, in effect, measures the amount of roughness in the fitted model.
- Define a smoothing parameter, λ, such that if $\lambda = 0$, there is no penalty for increased roughness, but as λ increases, roughness is penalised more and more.
- Estimate the parameters of the model, including the number of splines, by maximising the penalised log likelihood:

  $$l_p(\theta) = l(\theta) - \frac{1}{2}\lambda P(\theta),$$

  where $l(\theta)$ is the log likelihood from model

The penalised log likelihood is effectively trying to balance smoothness and adherence to the data.

4.1 Forecasting using p-Splines

Forecasting using p-splines is effected at the same time as the fitting of the model to past data. The past data used will consist of deaths and exposures at ages x for a range of past years t.

Forecasting may be carried out for each age separately, or for many ages simultaneously. In the case of a single age x, we wish to construct a model of the time series of mortality rates $m_{x,t}$, for age x over a period of years, so the knots are specified in terms of years.

Having decided upon the forecasting period (the number of years into the future we wish to forecast mortality), we add to the data set dummy deaths and exposures for each year in the forecast period. These are given a weight of 0 when estimating the model, whereas the existing data are given a weight of 1. This means that the dummy data have no impact on l(θ). We then choose the regression coefficients for the model so that the penalty P(θ) is unchanged.

4.1 Advantages & Disadvantages of p-Splines

The p-spline approach has the **advantages** that it is a natural extension of methods of graduation and smoothing, and it is relatively straightforward to implement in R.

It has the following **disadvantages:**

- When applied to ages separately, mortality at different ages is forecast independently so there is a danger that there will be roughness between adjacent ages. This can be overcome by fitting the model and forecasting in two dimensions (age and time) simultaneously.
- There is no explanatory element to the projection (in the way that time-series methods use a structure for mortality and an identifiable time series for projection).
- p-splines tend to be over-responsive to an extra year of data (though this can be ameliorated by increasing the knot spacing)

Question

CS2A April 2022

(i) Write down the two-factor Lee–Carter model, clearly defining each of the terms you use. [3]

A statistician is using the two-factor Lee–Carter model to project future mortality rates and has fitted the model to a set of mortality data. The statistician has observed that the fitted forces of mortality do not vary regularly with each calendar year but vary more regularly with age.

Therefore, the statistician suggests that, before projecting future mortality rates, the k parameters in the Lee–Carter model should be smoothed using penalised regression splines, but the a and b parameters should be kept as separate, unsmoothed parameters for each age.

(ii) Discuss the statistician's suggestion by considering each of the three parameters of the Lee–Carter model, k, a and b, in turn. [3]
[Total 6]

Solution

(i)

The two-factor Lee-Carter model may be written:

ln $m(x,t) = a\_x + b\_x * k\_t + epsilon(x, t)$ [1]

where:

$m(x,t)$ is the central mortality rate at age x in year t [1]

a_x describes the general shape of mortality at age x / a_x is the mean of the time-averaged logarithms of the central mortality rate at age x [½]

b_x measures the change in the rates in response to an underlying time trend in the level of mortality of k_t. [½]

$epsilon(x, t)$ are independently distributed normal random variables with means of zero and some variance to be estimated. [½]

[Marks available 3½, maximum 3]

Solution

(ii)

k parameters

The lack of regularity by calendar year is likely to be a genuine feature of the underlying process [1]

Mortality is subject to fluctuations from year to year, due to factors such as epidemics and harsh winters [½]

If the *k* parameters are smoothed using penalised spline regression and the penalty is used to project future mortality, then the projection is likely to place too much weight on the last few years of data [½]

a parameters

The *a* parameters represent the general shape of the mortality rates, which would be expected to vary in a regular manner with age [½]

This would suggest that smoothing might be required [½]

However, in this case, the fact that the fitted mortality rates vary regularly with age suggests that the *a* parameters also vary regularly enough to obviate the need for smoothing [½]

Smoothing of the *a* parameters would be more likely to be necessary for a small data set [½]

Solution

b parameters

The *b* parameters represent the variation of mortality improvements with age, which would be expected to be regular [½]

This would suggest that smoothing might be required [½]

Although the fitted mortality rates vary regularly with age in this case, it cannot be assumed that the *b* parameters also vary regularly enough to obviate the need for smoothing [½]

The *b* parameters require a higher volume of data to estimate reliably than the *a* parameters, increasing the likely need for smoothing [½]

If the *b* parameters are not smoothed, then the projected mortality rates at successive ages may cross over in the later years of the projection, which is not intuitively reasonable [½]

[Marks available 6½, maximum 3]

5 Methods based on Explanation

- The previous approaches take no, or only limited, cognisance of the causal factors underlying mortality. Since causal factors are quite well understood, at least at a general level, it might be thought sensible to use this knowledge in forecasting.
- For example, if cancer is a leading cause of death in a country, and if it seems likely that a significant breakthrough in the treatment of cancer is likely, this could be explicitly taken account of in mortality projections for that country. It might seem surprising that progress in this direction has been limited.
- However, the methods require either that long lags are specified between changes in the risk factors and changes in mortality, or that the risk factors themselves be forecasted, and forecasting developments in the risk factors is almost as difficult as forecasting mortality.

JAQS

5 Methods based on Explanation

Decomposition of mortality by cause of death is an integral part of the explanatory approach. However, in practice it is difficult to achieve successfully.

The reasons include:

- cause of death reporting is unreliable, especially at older ages (where an increasing proportion of deaths occur);
- causes of death often act synergistically, so it is not realistic to posit a single cause of death;
- elimination of one cause of death might "unmask" another cause that would not have been identified previously;
- the time series of data are often rather short.

Smoker [illegible]

cancer time

age

6 Sources of error in mortality forecasts

Mortality forecasts are always wrong. It is of interest to know how wrong they are likely to be, and what the main sources of error are. The latter is important not so much because it will help to eliminate errors (this is not possible) but so effort can be focused on the areas most likely to cause the forecasts to be at variance with reality, or on the elements of the process to which the sensitivity of the outcome is greatest.

Alho, J.M. (1990) 'Stochastic methods in population forecasting', International Journal of Forecasting 6, pp. 521-30, classified sources of error as follows:

1. Model mis-specification. We might have the wrong parameterisation function, or the wrong model.
2. Uncertainty in parameter estimates.
3. Incorrect judgement or prior knowledge. The data set we use as the basis may not accurately reflect the mortality we wish to model.
4. Random variability, including the randomness in the process generating the mortality, short term variation due to severe winters or hot summers.
5. Errors in data (for example age-misstatement).

back testing

in-sample bt

out-sample bt

Question

CS2A September 2019 Q2

The government of a small country is interested in projecting mortality over the next 20 years. Life tables are available by single years of age for each calendar year for the past 10 years. The country has a new Chief Statistician who suggests fitting an exponential curve to the time trend in mortality at each age x and forecasting mortality separately at each age using the parameters estimated for that age.

(i) Comment on this suggestion. [3]
(ii) Suggest an alternative approach which may be more suitable for projecting mortality in this country. [1] [Total 4]

Solution

2(i)

- The use of an exponential curve is attractive as there is evidence that age-specific mortality has declined exponentially in some past periods. [½]
- The approach is simple to understand and easy to implement. [1]
- However, fitting separate curves at each age risks the projected future mortality rates in any given year not progressing smoothly with age (and even decreasing with age in age ranges where this is implausible) i.e. under-graduated rates [1]
- This problem could be overcome by graduating the projected rates. [½]
- Or by using an alternative method/model in the first place. [½]
- The approach assumes that developments in medical technology, lifestyle, etc. in the future will progress steadily as they have in the past 10 years. [1]
- The appropriateness of this projection method may depend on whether the past history displays an exponential change over time [½]
- Using cohorts to project mortality instead of time period may lead to improvements in the reliability of the projection [½]
- It could be argued that 10 years of historic Life tables may not be sufficient to provide a reliable projection of future mortality [½]

[6, Max 3]

Solution

(ii)

EITHER: Lee-Carter model
OR: age, cohort model
OR: age, period, cohort model
OR: penalised splines
OR: decomposition of mortality by cause of death / Explanatory method. [1]

OR:
Adjust rates using projected rates from a similar country [½]

[1½, Max 1]

[Total 4]