Lecture

INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

**Class:** SY BSc

**Chapter Name:** Mortality projection

# Topics to be covered

1. Mortality projection – What & why?

2. Types of methods

    1. Based on expectations
    2. Based on extrapolation
        1. Lee Carter model
        2. Age period cohort model
        3. P-splines method
    3. Based on explanation

3. Sources of error in mortality forecasts

# Mortality projection

Mortality projection – What & why?

# Based on expectation

- Predicts mortality based on simple expectations
- Ex. – Continued exponential decline in mortality rates
- Mortality targets are set – either fitting deterministic functions or using expert opinions
- Mortality rates for person aged ‘x’ in a future year ‘t’ is given by:
    - $q_{x,t} = q_{x,0} \text{ X } R_{x,t}$
- $R_{x,t}$ – measures proportion by which mortality rate is expected to be reduced by future year ‘t’
- $R_{x,t} = \alpha_x + (1 - \alpha_x)(1 - f_{n,x})^{t/n}$
- $\alpha_{x,t}$ – Ultimate reduction factor
- $f_{n,x}$ – proportion of total decline expected to occur in ‘n’ years

# Based on expectation

- Special cases:
- $t = 0$
- $t = n$
- $t \rightarrow \infty$

# Based on expectation

- Example-

| Age | $\alpha_x$ | $f_{20,x}$ |
|---|---|---|
| 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

# Based on expectation

- Advantages

- Easy and straightforward
- Suitable for short term
- Follows simple deterministic approach

- Disadvantages

- Not suitable for population where there is rapid change in mortality
- Target setting leads to underestimation of true level of uncertainty around the forecast (ex. - smoking related deaths)

# Based on extrapolation

- Methods based on stochastic forecast models

- 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

# Based on extrapolation

1. **Lee Carter model**

- 2 factor model – age and period
- Formula – $\ln m_{x,t} = a_x + b_x k_t + \xi_{x,t}$
- $m_{x,t}$ – central rate of mortality of age 'x' in year 't'
- $a_x$ - general shape of mortality at age 'x'
- $b_x$ – measures change in rates in response to underlying time trend $k_t$
- $k_t$ – effect of time trend on mortality at time 't'
- $\xi_{x,t}$ – iid rv's with mean 0 and some variance

# Based on extrapolation

1. **Lee Carter model**

- In order to estimate mortality rates, we ignore error term and put some constraints
- $$\ln m_{x,t} = a_x + b_x k_t$$

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

- Usual constraints:
- 1)
- 2)

# Based on extrapolation

1. **Lee Carter model**

- Key points:
- $a_x$ =

- $b_x$ (example – age 60 – 0.24 and age 65 – 0.22) -

- $k_t$ (example – 1.5 – 0.075t) -

# Based on extrapolation

1. **Lee Carter model**

- Estimation of parameters:

- Approach 1

1. $a_x$ calculated as before

2. $\ln m_{x,t} - a_x = b_x k_t + \xi_{x,t}$

   We use singular value decomposition technique to find best combination of $b_x$ and $k_t$ values that fit the observed value $\ln m^{\wedge}_{x,t} - a^{\wedge}_{x}$

# Based on extrapolation

**1. <u>Lee Carter model</u>**

- Estimation of parameters:

- <u>Approach 2 (Macdonald et al)</u>

1. Method is based on generalized linear model

2. Parameters $a_x$, $b_x$ and $k_t$ are estimated using '*gnm*' function in R

3. Simple adjustment is made to the parameters estimated so that they satisfy the constraints

# Based on extrapolation

1. **<u>Lee Carter model</u>**

- <u>Forecasting using this model</u>

- 2 age parameters and 1 time parameter

- $a_x$ and $b_x$ are held constants, and $k_t$ is forecasted in the model

- $k_t$ can be modelled using Random walk series i.e. $k_t - k_{t-1} = \mu + \varepsilon_t$

    - Where $\mu$ measures avg. change in $k_t$
    - $\varepsilon$ are iid rv's with mean 0 and variance $\sigma^2$

- If we have data for year 't', we can forecast $k_{t+1}$ by:

$$\hat{k}_{t+1} = \hat{k}_t + \hat{\mu}$$

$$\hat{k}_{t+l} = \hat{k}_t + l\,\hat{\mu}$$

# Based on extrapolation

1. **Lee Carter model**

- Forecasting using this model

- Variance – As we assume a random walk model for $k_t$, uncertainty in the parameter keeps increasing with time i.e. error term keeps accumulating over time.

- Var($\mu$^) =

- Var($k^{\wedge}_{t+l}$) =

- Future value :

- Predicted future mortality:

# Based on extrapolation

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

# Based on extrapolation

**2. Age period cohort model**

- Lee Carter model with a cohort factor
- Formula - $\ln m_{x,t} = a_x + b^1_x k_t + b^2_x h_{t-x} + \xi_{x,t}$
- $b^1_x$
- $b^2_x$
- $h_{t-x}$

# Based on extrapolation

**2. Age period cohort model**

- Advantages
- Cohort effects are smaller but non negligible
- Models that take this into account have proven to be superior than 2 factor models of age and period

- Disadvantages
- There are heavy data requirements for cohort effects
- There can be an identification problem as there are 3 factors linearly dependent on each other

# Based on extrapolation

**3. Using p-splines**

- We can determine the no. of splines and degrees of polynomials in each spline to use them for forecasting using a linear regression model
- Under Gompertz model, ln $E(D_x)$ = ln $(E_x^c)$ + $\alpha + \beta_x$

- $E(D_x)$ – expected deaths at age 'x'

- If we replace $\alpha + \beta_x$ by $\sum_{j=1}^{s} O_j$ $B_j$ (x), we have

- ln $E(D_x)$ = ln $(E_x^c)$ + $\sum_{j=1}^{s} O_j$ $B_j$ (x)
- $B_j$ - set of splines, $O_j$ - parameters and S – no. of splines
- A relatively inflexible function is replaced by a more flexible one

- A penalised log-likelihood term may be used to create a balance between smoothness and adherence to data

# Based on extrapolation

**3. Using p-splines**

- Forecasting using p-splines
- Spline function can be used to model values of ln $m_{x,t}$ by t
- Formula - ln ($m_{x,t}$) = $\sum_{j=1}^{s} O_j$ $B_j$ (t)
- Forecasting is done for each age separately and knots are built based on time ‘t’ in years
- Package *MortalitySmooth* can be used in R to model this

# Based on extrapolation

**3. Using p-splines**

- Advantages

- It is a natural extension of methods of graduation and smoothing, and straightforward to implement in R

- Disadvantages

- There could be roughness between adjacent ages (as smoothing is done across 't' and not 'x'). This can be overcome by fitting model and forecasting in 2 dimensions (age and time)
- No natural interpretation to the final formulas and parameters
- It is over-responsive to an extra year of data

# Based on explanation

- Causes of factors are taken into account to understand changes in mortality

- Ex. – changes in death rates due to cancer/ heart diseases after invention of a new treatment

Q. Suppose the current mortality rate at age 70 for males in a particular country is 0.01 pa. An analysis of the causes of mortality at this age reveals that 35% of deaths are due to heart disease, 40% from cancer, and 25% from all other causes.
Due to the introduction of a revolutionary new treatment, next year it is predicted that the mortality rate of males aged 70 due to cancer will be 90% of the current rate. At the same time, it is predicted that the mortality rate due to heart disease will increase by 2% and that due to other causes will increase by 1% of their current rates.

i) Give a possible reason why the deaths from causes other than cancer might have increased.

ii) Calculate the expected population mortality rate for male lives aged 70 in one year's time.

- Multiple state Markov model may be used to achieve greater sophistication in methods based on explanation

# Sources of errors in mortality forecasts

- Mortality forecasts are always wrong

- Some common sources of errors are:
    - Model mis-specification. We might have the wrong parameterisation function, or the wrong model.
    - Uncertainty in parameter estimates.
    - Incorrect judgement or prior knowledge
    - Errors in data (for example age-misstatement).

- Forecast errors are usually positively correlated across ages or across times, and hence, they reinforce these into one another to widen the prediction interval