Projecting mortality rates using a Markov chain

Jaap Spreeuw

Joint work with:

Iqbal Owadally and Muhammad Kashif

Bayes Business School (formerly Cass), City, University of London

Longevity 18 Conference, London, UK, September 2023

Outline of contents

- Worked example as in Norberg (2013)
- Aim of this paper
- Model
- Calibration to data
- Forecasting
- Applications in life insurance and pensions
- Avenues for future research

Worked example as in Norberg (2013)

- Stochastic mortality changes modelled through time-homogeneous Markov chain
- Three causes of death:
  - First two causes are in force at all ages until disappearing independently at random times which are exponentially distributed.
  - Third cause remains in force, alternating between low and high mortality.

Aim of this paper

- Calibrate suitable time-homogeneous Markov chain model of stochastic mortality to real data.
- Challenges:
  - Cause-specific mortality: possible correlation between causes of death.
  - Mortality in aggregate terms: usually only *number of deaths* and *risk exposure* observed.

Model

- A Markov chain with $N+1$ "Alive" states and one "Dead" state.
- For individual aged x, transition of one "Alive" state i to next one, $i+1$, rate $\lambda(i)$, leads to change in mortality of $\exp\left[b_x \gamma(i+1)\right]$.
- From last "Alive" state only transition to death state possible.

Model

- "Alive" state i for life aged x $\left(i \in \{0,\ldots,N-1\}\right)$

State *i*

$$\mu_x^{(i)} = \left\{\prod_{j=0}^{i} \exp\left[b_x \gamma(j)\right]\right\} \cdot (\text{average mortality at age } x)$$

$\lambda(i)$

$\mu_x^{(i)}$

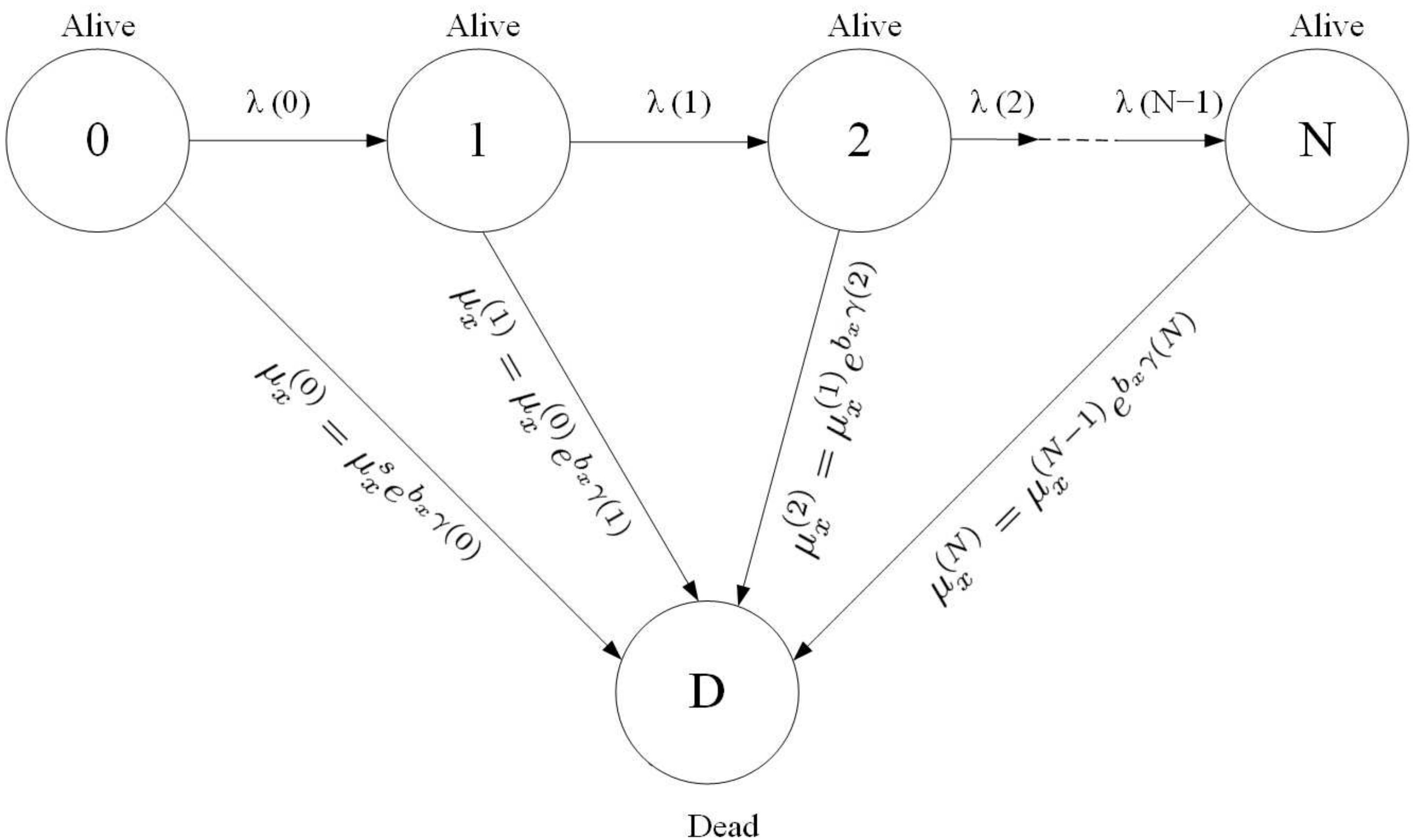
State *i* +1

$$\mu_x^{(i+1)} = \left\{\prod_{j=0}^{i+1} \exp\left[b_x \gamma(j)\right]\right\} \cdot (\text{aver. mort. at age } x)$$

Dead

Model

- Complete transition diagram:



Model; data

- Simpler model (*Model* 1): ignore age effects (so $b_x \equiv 1$).
- Data from Human Mortality Database: UK females, ages 20 to 104.
  - In sample: years Y to 2000;
  - Out-of-sample: years 2001 to 2016.
- For various N, find $\lambda(i), i \in \{0,\ldots,N-1\}$, and $\gamma(i), i \in \{0,\ldots,N\}$, by minimizing total weighted average quadratic distance (WAQD) between expected mortality and observed mortality.

Calibration to data

- Then year t has a contribution to the total WAQD of:

$$\sum_{k=0}^{N} p_{0k}(t) \sum_{x=20}^{104} w_{x,t} \left(b_x \sum_{j=0}^{k} \gamma(j) + a_x - \ln \hat{\mu}_{x,t} \right)^2$$

$w_{x,t}$: weight assigned to age x in year t

$\hat{\mu}_{x,t}$: empirical mortality rate age x in year t

a_x : weighted average log-mortality for age x

$p_{0k}(t)$: transition probability "Alive" state k mid year t

Calibration to data

- First, consider simple model.
- Ideally, the $2N+1$ parameters $\lambda(i), i \in \{0,\ldots,N-1\}$, and $\gamma(i), i \in \{0,\ldots,N\}$, found simultaneously by minimizing WAQD using standard calculus.
  - Problem: not feasible or very time consuming except for small N.
- Pragmatic alternative: fix values $\{\lambda(k)\}_0^{N-1}$ and derive optimal estimates of $\{\gamma(k)\}_0^{N}$. Then perform grid search for minimum WAQD on grid spanned by $\{\lambda(k)\}_0^{N-1}$.

Calibration to data

- Define $\Gamma(k) = \sum_{j=0}^{k} \gamma(j), k \in \{0,\dots,N\}$. Optimal values:

$$\hat{\Gamma}(k) = \frac{\sum_{t=1}^{51} p_{0k}(t) \sum_{x=20}^{104} w_{x,t} \left(a_x - \ln \hat{\mu}_{x,t}\right)}{\sum_{t=1}^{51} p_{0k}(t) \sum_{x=20}^{104} w_{x,t}}$$

- Note the interpretation!

Calibration to data

- Assume λ to be constant, so $\lambda = \lambda(0) = \lambda(1) = \ldots$
- The greater N:
  - the smaller the minimum WAQD.
    - Interpretation: ......
  - The greater λ for which WAQD is minimized.
    - Interpretation: ....
  - The closer $\gamma(i), i \in \{1,\ldots,N\}$, are to 1.
    - Interpretation:...

Calibration to data

- Comprehensive model with b_x (*Model* 2):
  - Starting values: optimal $\{\gamma(k)\}_0^N$ from simple model as before, and $b_x \equiv 1$.
  - Find optimal $\{\gamma(k)\}_0^N$ and $\{b_x\}_{20}^{104}$ by successive substitution.

Calibration to data

- Model 2 applied with $N = 50$:
  - We estimate $\hat{\lambda} = 1.29$.
  - Pattern of $\Gamma(k) = \sum_{j=0}^{k} \gamma(j), \quad k \in \{0,\dots,N\}$:
    - Strongly positive for 0;
    - Decreasing as function of state, becoming negative upon reaching state 33 (and more negative thereafter).
  - Pattern of $\{b_x\}_{20}^{104}$:
    - Relatively high for young ages, compared to later ages;
    - Relatively small for very high ages.

Forecasting

- Markov model augmented by new "Alive" states.
- Issue: standard time series methods do not seem to work.
- Instead: employ innovations state space model from Hyndman et al. (2008) to forecast Γ or γ.

Forecasting

- Innovations state space model:
  - Observation equation:

$$\gamma(k) = l_{k-1} + \phi b_{k-1} + \varepsilon_k$$

  - State equations:

$$l_k = l_{k-1} + \phi b_{k-1} + \alpha \varepsilon_k$$

$$b_k = \phi b_{k-1} + \beta \varepsilon_k$$

l_k : level of data, superposed on trend b_k, along with additive noise ε_k.

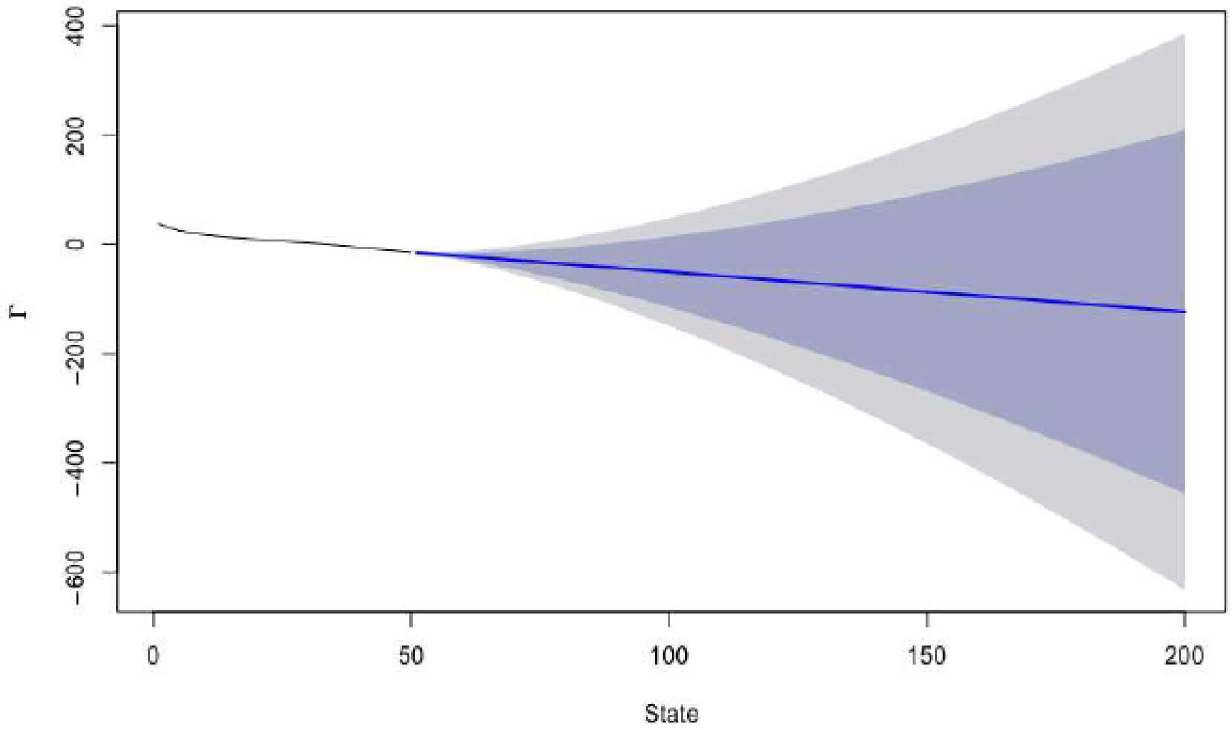
$$\varepsilon_k \sim N\left(0, \sigma_\varepsilon^2\right)$$

Forecasting

- Innovations state space model:
  - Last value of $\gamma(k)$ or $\Gamma(k)$ omitted.
  - Best model selected by bias corrected AIC.
  - Parameter values estimated by ML.
- Case considered:
  - $w_{x,t} \equiv 1$

Forecasting

- Illustrative plot of $\Gamma(k)$ for Model 2 with $N = 50$ with forecasts and confidence intervals.



Forecasting

- Forecast error for an out-of-sample year t:

$$\sum_{x=20}^{104}\left[\left(b_x\sum_{k=0}^{N}p_{0k}(t)\sum_{j=0}^{k}\gamma(j)\right)+a_x-\ln\hat{\mu}_{x,t}\right]^2$$

- Three model considered:
  - Naïve model with static mortality as in 2000;
  - Model 2 with $N=50$ and $Y=1950$.
  - Model 2 with $N=10$ and $Y=1990$.

Forecasting

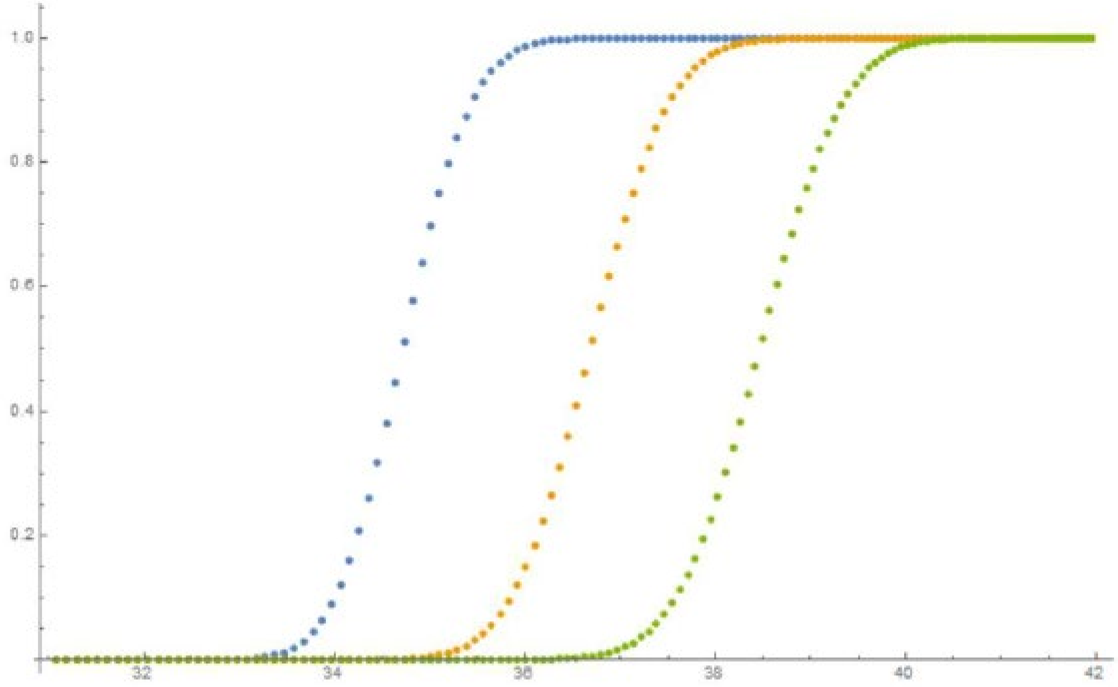
- Three model considered:
  - Naïve model with static mortality as in 2000.
  - Model 2 with $N = 50$ and $Y = 1950$.
  - Model 2 with $N = 10$ and $Y = 1990$.
- Observations:
  - In terms of <u>total forecast error</u>, Markov models outperform naïve model
  - Markov model with fewer states outperforms other Markov model (seems a bit surprising).

Applications in life insurance and pensions

- Common measures of mortality changes:
  - Distributions of expectations of life
  - Distributions of present values of annuities
- Markov model enables exact calculation of these measures by solving Thiele's differential equations.
- Examples concern
  - Durations 25, 40 and 55 (so calendar years 2000, 2015 and 2030), and
  - Ages 50 and 80.

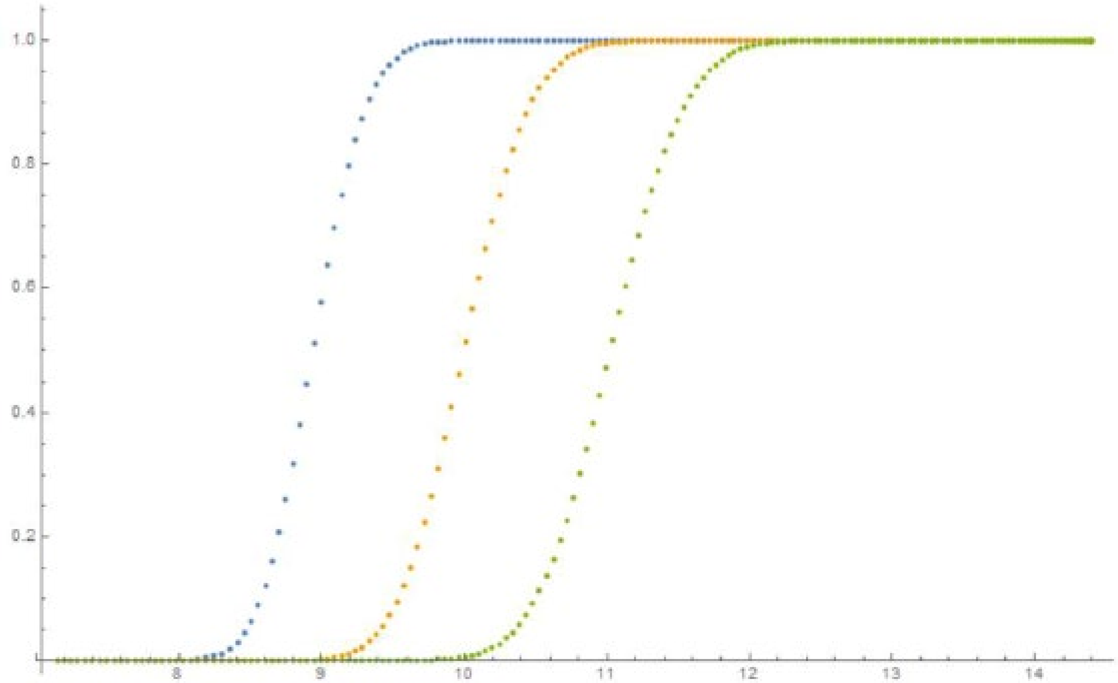
Applications in life insurance and pensions

- Cumulative distribution function of complete expectation of life for age 50:



Applications in life insurance and pensions

- Cumulative distribution function of complete expectation of life for age 80:



Avenues for future research

- Use criteria other than WAQD
- Relax assumption of constant transition intensities
- Allow for idiosyncratic shocks
- Forecasting:
  - More rigorous and systematic investigation into impact of combination of factors (number of states, period of investigation, transition intensity)
  - Use information on prediction intervals

References

- Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D. (2008). *Forecasting with exponential smoothing: the state space approach.* Springer, Berlin.
- Norberg, R. (2013). Optimal hedging of demographic risk in life insurance. *Finance and Stochastics* 17(1), 197-222.